National Congregations Study
Weights Documentation for NCS Waves I through IV
Last Edited: 23 July 2020

Stephanie Eckman, NORC at the University of Chicago
Steven Pedlow, NORC at the University of Chicago
Mark Chaves, Duke University

Key NCS Features Relevant for Weighting:

- Congregations enter the NCS sample by being named by GSS respondents as the place they attend religious services. Since larger congregations are more likely to be named than smaller congregations, the NCS is a probability-proportional-to-size sample. Using different weights, the data can be analyzed at the congregation level or at the attendee level.
- Waves I and II have a panel aspect. Waves III and IV also have a panel aspect.
- Wave III contained an oversample of congregations nominated by Hispanic GSS respondents.
- There are three types of weights: two that allow users to analyze data at the congregation level (one that ignores duplicate nominations and one that takes account of duplicate nominations) and one that allows users to analyze data at the attendee level.
    - If you are interested in the average congregation, then use the congregation-level weights.
    - If you are interested in the congregation of the average attendee, use the attendee-level weight.

SHORTCUT:

- **USE WT_ALL4_CONG_DUP TO DESCRIBE THE AVERAGE CONGREGATION IN EACH WAVE.**

- **USE WT_ALL4_ATTENDEE TO DESCRIBE THE CONGREGATION OF THE AVERAGE ATTENDEE IN EACH WAVE.**

FOR MOST PURPOSES THESE WILL BE THE ONLY WEIGHTS NEEDED.

This document contains:

Part 1. WHICH WEIGHTS ARE APPROPRIATE FOR DIFFERENT POPULATIONS/ANALYSES OF INTEREST?

a. Population of interest: All four waves of the NCS, using all cases from any NCS wave.

These three weights combine all the cases from all four waves, making each NCS wave representative of congregations in that year:

> **WT_ALL4_CONG_IGN**: Weight for all cases, including the panel and oversample cases, ignoring duplicate nominations. This weight allows users to analyze the data at the congregation level.
>
> **WT_ALL4_CONG_DUP**: Weight for all cases, including the panel and oversample cases, taking account of duplicate nominations. This weight allows users to analyze the data at the congregation level.
>
> **WT_ALL4_ATTENDEE**: Weight for all cases, including the panel and oversample cases, which allows users to analyze the data at the attendee level.

b. Analysis of Interest: All data from any single NCS-IV wave, combining the newly nominated, panel, and oversample cases.

The same weights as in (a) can be used with the condition (YEAR=1999, YEAR=2006, YEAR=2012, or YEAR=2018) to analyze data from any particular NCS wave: **WT_ALL4_CONG_IGN, WT_ALL4_CONG_DUP,** and **WT_ALL4_ATTENDEE**.

c. Population of Interest: Only congregations newly nominated in 2018

These three weights are restricted to the 744 congregations nominated in 2018 (SET = K, L, M, and N):

> **WT_WAVE4_CONG_IGN**: Weight for cases nominated in 2018, ignoring duplicate nominations. This weight allows users to analyze the data at the congregation level. A similar weight for NCS-II is WT_WAVE2_CONG_IGN.
>
> **WT_WAVE4_CONG_DUP**: Weight for cases nominated in 2018, taking account of duplicate nominations. This weight allows users to analyze the data at the congregation level. A similar weight for NCS-II is WT_WAVE2_CONG_DUP.
>
> **WT_WAVE4_ATTENDEE**: Weight for cases nominated in 2018, which allows users to analyze the data at the attendee level. A similar weight for NCS-II is WT_WAVE2_CONG_ATTENDEE.

These weights should be used by analysts who want to ignore panel cases when analyzing Wave IV data.

d. Population of Interest: The NCS-III/IV panel.

There were 1,331 2012-nominated congregations completed in NCS-III. The 830 NCS-III cases that that were nominated by GSS respondents who were initially sampled in 2012 (as opposed to the 2012 GSS respondents who were part of the 2008 and 2010 GSS panel samples) were invited to participate again in NCS-IV. Because of the structure of the GSS, these 830 constitute a representative random sample of U.S. congregations in 2012, and 597 of them participated in NCS-IV. An additional five 2018-nominated congregations also participated in the NCS-III but were not among the 830 NCS-III congregations eligible for the panel. Just as we included Set G in the NCS-I/II panel, we include Set N in the NCS-III/IV panel because we have data in both rounds for them. Therefore, we have 602 cases in the NCS-III/IV panel (SET = J, M, N).

If you want to use Wave III and Wave IV data for the 602 panel cases, three new weights have been created:

> **WT_PANEL34_CONG_IGN**: Weight for panel cases from Waves III and IV, ignoring duplicate nominations, which allows users to analyze the data at the congregation level.
>
> **WT_PANEL34_CONG_DUP**: Weight for panel cases from Waves III and IV, taking account of duplicate nominations. This weight allows users to analyze the data at the congregation level.
>
> **WT_PANEL34_ATTENDEE**: Weight for panel cases from Waves III and IV, which allows users to analyze the data at the attendee level.

These weights allow analysts to examine how a representative sample of congregations in 2012 changed between Waves III and IV.

There are 1,204 total cases with non-missing values on these weights: 602 with YEAR=2012 and 602 with YEAR=2018.

These are analogous to WT_PANEL12_CONG_IGN and WT_PANEL12_ATTENDEE for the Wave I/II panel. WT_PANEL12_CONG_DUP was not created for the Wave I/II panel.

e. Population of Interest: Only the NCS-III cases nominated by Hispanic GSS 2012 respondents

The NCS-III included a Hispanic oversample, so three special weights were created for analysts who want to examine only the 235 NCS-III congregations nominated by Hispanic GSS 2012 respondents. These weights make this set of congregations representative of congregations attended by Hispanics in 2012.

> **WT_WAVE3_HISP_CONG_IGN**: Weight for cases from Wave III nominated by a Hispanic GSS 2012 respondent, ignoring duplicate nominations, which allows users to analyze the data at the congregation level.
>
> **WT_ WAVE3_HISP_CONG_DUP**: Weight for cases from Wave III nominated by a Hispanic GSS 2012 respondent, taking account of duplicate nominations. This weight allows users to analyze the data at the congregation level.
>
> **WT_ WAVE3_HISP_ATTENDEE**: Weight for cases from Wave III nominated by a Hispanic GSS 2012 respondent, which allows users to analyze the data at the attendee level.

Congregations nominated by Hispanic GSS respondents in other NCS waves are not identified in the public NCS dataset.

f. Analysis of Interest: NCS-III cases excluding the Hispanic oversample

Three special weights were created to exclude the 77 congregations that entered through the Hispanic oversample in NCS-III:

> **WT_WAVE3_NOHISP_CONG_IGN**: Weight for cases from Wave III EXCLUDING those from the Hispanic oversample, ignoring duplicate nominations, which allows users to analyze the data at the congregation level.
>
> **WT_ WAVE3_NOHISP_CONG_DUP**: Weight for cases from Wave III EXCLUDING those from the Hispanic oversample, taking account of duplicate nominations. This weight allows users to analyze the data at the congregation level.
>
> **WT_ WAVE3_NOHISP_ATTENDEE**: Weight for cases from Wave III EXCLUDING those from the Hispanic oversample, which allows users to analyze the data at the attendee level.

Users who want to analyze all four waves EXCLUDING the Hispanic oversample can use the appropriate WT_ALL4* weight for Waves I, II, and IV (YEAR = 1998, 2006, 2018) while using the appropriate WT_WAVE3_NOHISP* weight for Wave III.

g. Population of Interest: Only congregations newly nominated in 2006

These three weights are restricted to the 1,254 congregations nominated in 2006:

> **WT_WAVE2_CONG_IGN**: Weight for cases nominated in 2006, ignoring duplicate nominations. This weight allows users to analyze the data at the congregation level.
>
> **WT_WAVE2_CONG_DUP**: Weight for cases nominated in 2006, taking account of duplicate nominations. This weight allows users to analyze the data at the congregation level.
>
> **WT_WAVE2_ATTENDEE**: Weight for cases nominated in 2006, which allows users to analyze the data at the attendee level.

These weights should be used by analysts who want to ignore panel cases when analyzing Wave II data.

h. Population of Interest: The NCS-I/II panel

There were 1,234 1998-nominated congregations in NCS-I. A randomly selected 325 of these congregations were invited to participate again in NCS-II, and 256 of these did so. An additional six 2012-nominated congregations also participated in the NCS-I but were not among the 325 NCS-I congregations selected for the panel. Nevertheless, we included these six congregations in the panel because we have data in both rounds for them. Therefore, we have 262 cases in the NCS-I/II panel (SET = C, F, G).

If you want to use all Wave I and Wave II data for the 262 panel cases, two weights were created during Wave II:

> **WT_PANEL12_CONG_IGN**: Weight for panel cases from Waves I and II, ignoring duplicate nominations, which allows users to analyze the data at the congregation level.
>
> **WT_PANEL12_ATTENDEE**: Weight for panel cases from Waves I and II, which allows users to analyze the data at the attendee level.
>
> (**WT_PANEL12_CONG_DUP** was not created.)

These weights allow analysts to examine how a representative sample of congregations in 1998 changed between Waves I and II.

There are 524 total cases with non-missing values on these weights: 262 with YEAR=1998 and 262 with YEAR=2006.

Part 3. WEIGHTING VARIABLES (YEAR, SET, PANEL)

There are fourteen sets of completed cases:

Year	Set	Description	Number of Cases
1998	A	1998 data for 1998-nominated congregations that were not selected for the panel.	909
1998	B	1998 data for 1998-nominated congregations that were randomly selected for the panel.	325
2006	C	2006 data for the 1998-nominated congregations that were randomly selected for the panel and not re-nominated in 2006	252
2006	D	2006 data for the 2006-nominated congregations established before or during 1998.	1,194
2006	E	2006 data for the 2006-nominated congregations established after 1998.	50
2006	F	2006 data for the 2006-nominated congregations that were also in the 1998 sample and were randomly selected for the panel.	4
2006	G	2006 data for the 2006-nominated congregations that were also in the 1998 sample but were not randomly selected for the panel	6
2012	H	2012 data for the 2012-nominated congregations not eligible for the panel (like A)	501
2012	I	2012 data for the 2012-nominated congregations that were eligible for the panel (like B)	830
2018	J	2018 data for the 2012-nominated congregations that were selected for the panel and not re-nominated in 2018 (like C)	518
2018	K	2018 data for the 2018-nominated congregations established before or during 2012 (like D)	637
2018	L	2018 data for the 2018-nominated congregations established after 2012 (like E).	23
2018	M	2018 data for the 2018-nominated congregations that were also in the 2012 sample and were eligible for the panel (like F).	79
2018	N	2018 data for the 2018-nominated congregations that were also in the 2012 sample but were not eligible for the panel (like G)	5

For each case in these sets, there are twenty-three variables relevant to weighting:

a. YEAR: variable indicating year of data collection:

Year	Number of Cases
1998	1,234
2006	1,506
2012	1,331
2018	1,262

b. PANEL: dummy variable indicating whether the case was part of a panel survey or not
   i. If SET='A', 'D', 'E', 'H', 'K', or 'L' PANEL=0
   ii. If SET='B', 'C', 'F', 'G', 'I', 'J', 'M', or 'N' PANEL=1

c. SET: character variable taking values A – N, as given above

d. The twenty weight variables listed above and below.

Part 3. DETAILS OF THE WEIGHTS

For each weight variable, we outline the calculation steps, using the following notation: $Wi_{<step>}^{<set>}$

where: *i* refers to the weight variable being calculated
set refers to the set of cases (A-N as given above)
step refers to the step in the calculation (0, 1… final)

WT_ALL4_CONG_IGN: Weight for all 5,333 cases from all four waves, including the panel and oversample cases, ignoring duplicate nominations. This weight allows users to analyze the data at the congregation level. We use the shorthand W1 for WT_ALL4_CONG_IGN in the formulas below.

WT_ALL3_CONG_IGN is copied into WT_ALL4_CONG_IGN for the 4,071 cases in NCS-I, NCS-II, and NCS-III. This weight is positive for all SET values (A-N). Here are the steps followed again in NCS-IV:

1. Calculate baseweights for all congregations nominated by GSS 2012 or GSS 2018 respondents (if nominated in GSS 2018, we use ONLY nominations from 2018).

   The 2012 and 2018 GSS designs included subsampling of households, so not all households have the same weight. To approximate what was done above, we use the minimum of the weights of the nominating households in the numerator. For example, if a congregation is nominated by both a subsampled household and a non-subsampled household, this weight counts the congregation as nominated by the non-subsampled household.

$$W1_0^J = \frac{\min_{i \subset c}(W_i^{2012})}{S_c}$$

$$W1_0^{K,L,M,N} = \frac{\min_{i \subset c}(W_i^{2018})}{S_c}$$

S_c is the size of the congregation as reported by the congregation itself: variable NUMADLTS in the 2018 data file.[1]

2. Let α be the sample estimate of the percent of congregations that were established after 1998, calculated using weighting sets K, L, M and N by $W1_0$. We use this proportion as a quality check in a later step.

$$\alpha = \frac{\sum W1_0^L}{\sum W1_0^J + \sum W1_0^K + \sum W1_0^L + \sum W1_0^M + \sum W1_0^N}$$

3. Identify congregations in the 2018 sample that were established after 2012 nominations and set aside. These congregations had no chance of selection in the previous round and must stand in for all new congregations.

   The final weight of the cases in set L is equal to the baseweight.

$$W1_2^L = W1_0^L$$

4. The 2012 panel cases and the 2018 cases that were established before 2012 are each a national probability sample of older congregations. To combine these two samples, we developed a trade-off parameter:

$$W1_1^{K,M,N} = \lambda * W1_0^{K,M,N}$$
$$W1_1^J = (1 - \lambda) * W1_0^J$$

   We calculated the optimal lambda which equalizes the contributions to effective sample size from each sample. See Appendix B for details on the derivation of the optimal lambda.

5. We next reduced the weights of set J, K, M and N cases so that their sum is equal to the weighted number of older cases recruited from the 2012 respondents.

$$W1_2^J = \gamma * W1_1^J$$
$$W1_2^{K,M,N} = \gamma * W1_1^{K,M,N}$$

[1] See Appendix A for details on how this variable was imputed when it was missing.

where $\gamma = \frac{\sum W1_1^K + \sum W1_1^M + \sum W1_1^N}{\sum W1_0^J + \sum W1_1^K + \sum W1_1^M + \sum W1_1^N}$

This adjustment is necessary so that when data from the older congregations in this round are combined with data from the newer congregations (set L from step 3), the weighted percent of new congregations in the combined sample equals the weighted estimate of the percent of new congregations in the population, α. We checked that α* = α.

$$\alpha^* = \frac{\sum W1_2^L}{\sum W1_2^J + \sum W1_2^K + \sum W1_2^L + \sum W1_2^M + \sum W1_2^M}$$

6. Rescale sets J, K, L, M, and N

   Many data analysis programs assume that the sum of the weights is equal to the sample size. Thus it is good practice to rescale the weights to the total number of cases, to ensure correct calculation of standard errors and confidence intervals. Without changing the relative weights between the cases, we rescaled the weights for sets J, K, L and M so that the sum of the weights is equal to the number of cases.

$$\beta = \frac{|J| + |K| + |L| + |M| + |N|}{\sum W1_2^J + \sum W1_2^K + \sum W1_2^L + \sum W1_2^M + \sum W1_2^N}$$

$$W1_{final}^{J,K,L,M,N} = \beta * W1_2^{J,K,L,M,N}$$

WT_ALL4_CONG_DUP: Weight for all 5,333 cases from all four waves, including the panel and oversample cases, taking account of duplicate nominations. This weight allows users to analyze the data at the congregation level. We use the shorthand W2 for WT_ALL4_CONG_DUP in the formulas below.

WT_ALL3_CONG_DUP is copied into WT_ALL4_CONG_DUP for the 4,071 cases in NCS-I, NCS-II, and NCS-III. This weight is positive for all SET values (A-N). Here are the steps followed again in NCS-IV:

1. Calculate baseweights for all congregations nominated by GSS 2012 or GSS 2018 respondents (if nominated in GSS 2018, we use ONLY nominations from 2018)

   The 2012 and 2018 GSS designs included subsampling of households, so the sum of the weights of the nominating respondents cannot be ignored.

$$W2_0^J = \frac{\sum_{i \subset c} W_i^{2012}}{S_c}$$

$$W2_0^{K,L,M,N} = \frac{\sum_{i \subset c} W_i^{2018}}{S_c}$$

The numerator sums the weights of all GSS respondents who nominated a congregation.

Steps 2 – 6 are unchanged from WT_ALL4_CONG_IGN. All parameters (α,α*,β,γ,φ,λ) were recalculated for WT_ALL4_CONG_DUP.

WT_ALL4_ATTENDEE: Weight for all 5,333 cases from all four waves, including the panel cases, which allows users to analyze the data at the attendee-level. We use the shorthand W3 for WT_ALL4_ATTENDEE in the formulas below.

WT_ALL3_ATTENDEE is copied into WT_ALL4_ATTENDEE for the 4,071 cases in NCS-I, NCS-II, and NCS-III. This weight is positive for all SET values (A-N). Here are the steps followed again in NCS-IV:

1. Calculate baseweights for all congregations nominated by GSS 2012 or GSS 2018 respondents (if nominated in GSS 2018, we use ONLY nominations from 2018)

   The 2012 and 2018 GSS designs included subsampling of households, so the sum of the weights of the nominating respondents cannot be ignored.

$$W3_0^{J} = \sum_{i \subset c} W_i^{2012}$$

$$W3_0^{K,L,M,N} = \sum_{i \subset c} W_i^{2018}$$

   The attendee weight sums the weights of all GSS respondents who nominated a congregation.

Steps 2 – 6 are unchanged from WT_ALL4_CONG_IGN. All parameters (α,α*,β,γ,φ,λ) were recalculated for WT_ALL4_ATTENDEE.

WT_WAVE4_CONG_IGN: Weight for 744 cases nominated in the 2018 GSS (SET=K, L, M, N), ignoring duplicate nominations. This weight allows users to analyze the data at the congregation level.

WT_WAVE4_CONG_IGN will be used by researchers who wish to analyze the 2018 nominated congregations only. Note that only cases in sets K, L, M, and N have non-missing values of WT_WAVE4_CONG_IGN. We use the shorthand W4 for WT_WAVE4_CONG_IGN in the formulas below.

1. Calculate baseweights for all congregations nominated by GSS 2018 respondents.

   The 2018 GSS design included subsampling of households, so not all households have the same weight. Therefore, we use the minimum of the weights of the nominating households in the numerator. For example if a congregation is nominated by both a subsampled household and a non-subsampled household, this weight counts the congregation as nominated by the non-subsampled household.

$$W4_0^{K,L,M,N} = \frac{\min_{i \subset c}(W_i^{2018})}{S_c}$$

2. Rescale weights.

   Many data analysis programs assume that the sum of the weights is equal to the sample size. Thus it is good practice to rescale the weights to the total number of cases, to ensure correct calculation of standard errors and confidence intervals. Without changing the relative weights between the cases, we rescaled the weights for sets K, L and M so that the sum of the weights is equal to the number of cases.

$$\beta = \frac{|K| + |L| + |M| + |N|}{\sum W4_0^K + \sum W4_0^L + \sum W4_0^M + \sum W4_0^N}$$

$$W4_{final}^{K,L,M,N} = \beta * W4_0^{K,L,M,N}$$

WT_WAVE4_CONG_DUP: Weight for 744 cases nominated in the 2018 GSS (SET=K, L, M, N), taking account of duplicate nominations. This weight allows users to analyze the data at the congregation level.

WT_WAVE4_CONG_DUP will be used by researchers who wish to analyze the 2018 nominated congregations only. Note that only cases in sets K, L, and M have non-missing values of WT_WAVE4_CONG_DUP. We use the shorthand W5 for WT_WAVE4_CONG_DUP in the formulas below.

1. Calculate baseweights for all congregations nominated by GSS 2018 respondents:

The 2018 GSS design included subsampling of households, so the sum of the weights of the nominating respondents cannot be ignored.

$$W5_0^{K,L,M,N} = \frac{\sum_{i \subset c} W_i^{2018}}{S_c}$$

The numerator sums the weights of all GSS respondents who nominated a congregation.

2. Rescale weights.

   Many data analysis programs assume that the sum of the weights is equal to the sample size. Thus it is good practice to rescale the weights to the total number of cases, to ensure correct calculation of standard errors and confidence intervals. Without changing the relative weights between the cases, we rescaled the weights for sets K, L and M so that the sum of the weights is equal to the number of cases.

$$\beta = \frac{|K| + |L| + |M| + |N|}{\sum W5_0^K + \sum W5_0^L + \sum W5_0^M + \sum W5_0^N +}$$

$$W5_{final}^{K,L,M,N} = \beta * W5_0^{K,L,M,N}$$

WT_WAVE4_ATTENDEE: Weight for 744 cases nominated in the 2018 GSS (SET=K, L, M, N), allowing users to analyze the data at the attendee level.

WT_WAVE4_ ATTENDEE will be used by researchers who wish to analyze the 2018 nominated congregations only. Note that only cases in sets K, L, M, and N have non-missing values of WT_WAVE4_ATTENDEE. We use the shorthand W6 for WT_WAVE4_ATTENDEE in the formulas below.

1. Calculate baseweights for all congregations nominated by GSS 2018 respondents:

   The 2018 GSS design included subsampling of households, so the sum of the weights of the nominating respondents cannot be ignored.

$$W6_0^{K,L,M,N} = \sum_{i \subset c} W_i^{2018}$$

   The attendee weight sums the weights of all GSS respondents who nominated a congregation.

2. Rescale weights.

Many data analysis programs assume that the sum of the weights is equal to the sample size. Thus it is good practice to rescale the weights to the total number of cases, to ensure correct calculation of standard errors and confidence intervals. Without changing the relative weights between the cases, we rescaled the weights for sets K, L and M so that the sum of the weights is equal to the number of cases.

$$\beta = \frac{|K| + |L| + |M| + |N|}{\sum W6_0^K + \sum W6_0^L + \sum W6_0^M + \sum W6_0^N}$$

$$W6_{final}^{K,L,M,N} = \beta * W6_0^{K,L,M,N}$$

WT_PANEL34_CONG_IGN: Weight for 602 panel cases (SET = I, J, M, N), ignoring duplicate nominations, which allows users to analyze the data at the congregation-level.

WT_PANEL34_CONG_IGN will be used by researchers who wish to analyze the congregations nominated and interviewed in 2012 and 2018. Note that only cases in sets I, J, M, and N have non-missing values of WT_PANEL34_CONG_IGN. We use the shorthand W7 for WT_PANEL34_CONG_IGN in the formulas below.

1. Merge NCS-III completed cases with NCS-IV completed cases.

2. Keep only those NCS-III respondents who also completed the interview in NCS-IV (some may have been nominated again in 2018). These cases are in SET = J, M, N for 2018.

3. Calculate baseweights for all congregations nominated by GSS 2012 respondents (IGNORING nominations from 2018; this uses NCS-III data).

   The 2012 GSS design included subsampling of households, so not all households have the same weight. To approximate what was done above, we use the minimum of the weights of the nominating households in the numerator. For example, if a congregation is nominated by both a subsampled household and a non-subsampled household, this weight counts the congregation as nominated by the non-subsampled household.

$$W7_0^{J,M,N} = \frac{\min_{i \subset c}(W_i^{2012})}{S_c}$$

4. Rescale weights.

   We scale the weights here so that the sum of the weights is equal to the number of cases.

$$\beta = \frac{|J| + |M| + |N|}{\sum W7_0^J + \sum W7_0^M + \sum W7_0^N}$$

$$W7_{final}^{J,M,N} = \beta * W7_0^{J,M,N}$$

NOTE: This methodology is equivalent to starting with WT_ALL3_CONG_IGN for the panel cases and re-scaling to the sample size.

WT_PANEL34_CONG_DUP: Weight for 602 panel cases (SET = I, J, M, N), taking account of duplicate nominations, which allows users to analyze the data at the congregation-level.

WT_PANEL34_CONG_DUP will be used by researchers who wish to analyze the congregations nominated and interviewed in 2012 and 2018. Note that only cases in sets I, J, M, and N have non-missing values of WT_PANEL34_CONG_DUP. We use the shorthand W8 for WT_PANEL34_CONG_DUP in the formulas below.

1. Merge NCS-III completed cases with NCS-IV completed cases.

2. Keep only those NCS-III respondents who also completed the interview in NCS-IV (some may have been nominated again in 2018). These cases are in SET = J, M, N for 2018.

3. Calculate baseweights for all congregations nominated by GSS 2012 respondents (IGNORING nominations from 2018; this uses NCS-III data).

   The 2018 GSS design included subsampling of households, so the sum of the weights of the nominating respondents cannot be ignored.

$$W8_0^{J,M,N} = \frac{\sum_{i \subset c} W_i^{2012}}{S_c}$$

   The numerator sums the weights of all GSS respondents who nominated a congregation.

4. Rescale weights.

   We scale the weights here so that the sum of the weights is equal to the number of cases.

$$\beta = \frac{|J| + |M| + |N|}{\sum W8_0^J + \sum W8_0^M + \sum W8_0^N}$$

$$W8_{final}^{J,M,N} = \beta * W8_0^{J,M,N}$$

NOTE: This methodology is equivalent to starting with WT_ALL3_CONG_DUP for the panel cases and re-scaling to the sample size.

WT_PANEL34_ATTENDEE: Weight for 602 panel cases (SET = I, J, M, N), which allows users to analyze the data at the attendee-level.

WT_PANEL34_ATTENDEE will be used by researchers who wish to analyze the congregations nominated and interviewed in 2012 and 2018. Note that only cases in sets I, J, M, and N have non-missing values of WT_PANEL34_ATTENDEE. We use the shorthand W9 for WT_PANEL34_ATTENDEE in the formulas below.

1. Merge NCS-III completed cases with NCS-IV completed cases.

2. Keep only those NCS-III respondents who also completed the interview in NCS-IV (some may have been nominated again in 2018). These cases are in SET = J, M, N for 2018.

3. Calculate baseweights for all congregations nominated by GSS 2012 respondents (IGNORING nominations from 2018; this uses NCS-III data).

   The 2018 GSS design included subsampling of households, so the sum of the weights of the nominating respondents cannot be ignored.

   $$W9_0^{J,M,N} = \sum_{i \subset c} W_i^{2012}$$

   The attendee weight sums the weights of all GSS respondents who nominated a congregation.

4. Rescale weights.

   We scale the weights here so that the sum of the weights is equal to the number of cases.

   $$\beta = \frac{|J| + |M| + |N|}{\sum W9_0^J + \sum W9_0^M + \sum W9_0^N}$$

$$W9^{J,M,N}_{final} = \beta * W9^{J,M,N}_{0}$$

NOTE: This methodology is equivalent to starting with WT_ALL3_ATTENDEE for the panel cases and re-scaling to the sample size.

WT_WAVE3_HISP_CONG_IGN for Wave III (sets H and I) cases: Weight for Wave III Hispanic cases, ignoring duplicate nominations. This weight allows users to analyze the data at the congregation level. We use the shorthand W10 for WT_WAVE3_HISP_CONG_IGN in the formulas below.

1. Subset WT_ALL4_CONG_IGN to only Wave III Hispanic nominators (there are 268 nominators that nominated 235 unique congregations):

$$W10^{H,I}_{1} = W1^{H,I}_{final}, \text{ if } i \subset \text{ Wave III and HISP1=1}$$
$$W10^{H,I}_{1} = \text{missing, if otherwise}$$

2. Rescale weights

   As in step 6 for WT_ALL4_CONG_IGN above, we scaled the weights here so that the sum of the weights is equal to the number of cases (235).

WT_WAVE3_HISP_CONG_DUP for Wave III (sets H and I) cases: Weight for Wave III Hispanic cases, taking account of duplicate nominations. This weight allows users to analyze the data at the congregation level. We use the shorthand W11 for WT_WAVE3_HISP_CONG_DUP in the formulas below.

1. Subset WT_ALL4_CONG_DUP to only Wave III Hispanic nominators (there are 268 nominators that nominated 235 unique congregations):

$$W11^{H,I}_{1} = W2^{H,I}_{final}, \text{ if } i \subset \text{ Wave III and HISP1=1}$$
$$W11^{H,I}_{1} = \text{missing, if otherwise}$$

2. Rescale weights

   As in step 6 for WT_ALL4_CONG_IGN above, we scaled the weights here so that the sum of the weights is equal to the number of cases (235).

WT_WAVE3_HISP_ATTENDEE for Wave III (sets H and I) cases: Weight for Wave III Hispanic cases, ignoring duplicate nominations. This weight allows users to analyze the data at the congregation level. We use the shorthand W12 for WT_WAVE3_HISP_ATTENDEE in the formulas below.

1. Subset WT_ALL3_ATTENDEE to only Wave III Hispanic nominators (there are 268 nominators that nominated 235 unique congregations):

$$W12_1^{H,I} = W3_{final}^{H,I}, \text{ if } i \subset \text{ Wave III and HISP1=1}$$
$$W12_1^{H,I} = \text{missing, if otherwise}$$

2. Rescale weights

   As in step 6 for WT_ALL4_CONG_IGN above, we scaled the weights here so that the sum of the weights is equal to the number of cases (235).

WT_WAVE3_NOHISP_CONG_IGN for Wave III (sets H and I) cases: Weight for Wave III cases without the Hispanic oversample cases, ignoring duplicate nominations. This weight allows users to analyze the data at the congregation level. We use the shorthand W13 for WT_WAVE3_NOHISP_CONG_IGN in the formulas below.

Step 0a. Nominators in GSS 2012 can be split into Cross-Sectional (CX) nominators who first responded to GSS in 2012, Panel respondents from GSS 2010 (P10), and Panel respondents from GSS 2008 (P08). We need to adjust the weights because only Hispanic nominators were allowed for P08 cases or P10 cases with ballot=B:

$$W13_a^{H,I} = \text{missing, if } i \subset P08 \text{ or } (i \subset P10 \text{ and ballot=B})$$
$$W13_a^{H,I} = W_i^{2012} * \frac{\sum_{i \subset CX} W_i^{2012} + \sum_{i \subset P10} W_i^{2012} + \sum_{i \subset P0} W_i^{2012}}{\sum_{i \subset CX} W_i^{2012} + \sum_{i \subset P10, ballot=A,C} W_i^{2012}} \text{ otherwise}$$

The former cases (set to missing) are the Hispanic nominators from the sample types that excluded non-Hispanic nominators. Since Hispanics and non-Hispanics were treated equally outside these sample types, they all receive the save adjustment.

Step 0b. Adjust all P10 weights to account for the lack of nominators among panel respondents first responding in 2010 who were given ballot “B”. In this case, it is a random one-third of cases excluded from the panel respondents first responding in 2010. Therefore, the adjustment is simply 3/2 = 1.5:

$$W13_b^{H,I} = 1.5 * W13_a^{H,I} \text{ , if } i \subset \text{P10}$$

$$W13_b^{H,I} = W13_a^{H,I} \text{ ,otherwise}$$

Steps 1 and 2 (Cross-sectional weight steps). These steps are the same as for WT_WAVE4_CONG_IGN.

WT_WAVE3_NOHISP_CONG_DUP for Wave III (sets H and I) cases: Weight for Wave III cases without the Hispanic oversample cases, taking account of duplicate nominations. This weight allows users to analyze the data at the congregation level.

Steps 0a and 0b (Subsampling of GSS respondents steps). These steps are the same as for WT_WAVE3_NOHISP_CONG_IGN.

Steps 1 and 2 (Cross-sectional weight steps). These steps are the same as for WT_WAVE4_CONG_DUP.

WT_WAVE3_NOHISP_ATTENDEE for Wave III (sets H and I) cases: Weight for Wave III cases without the Hispanic oversample cases, which allows users to analyze the data at the attendee-level.

Steps 0a and 0b (Subsampling of GSS respondents steps). These steps are the same as for WT_WAVE3_NOHISP_CONG_IGN.

Steps 1 and 2 (Cross-sectional weight steps). These steps are the same as for WT_WAVE4_ATTENDEE.

WT_WAVE2_CONG_IGN: Weight for 1,254 cases nominated in the 2006 GSS (SET=D, E, F, G), ignoring duplicate nominations. This weight allows users to analyze the data at the congregation level.

WT_WAVE2_CONG_IGN can be used by researchers who wish to analyze the 2006 nominated congregations only. Note that only cases in sets D, E, F, and G have non-missing values of WT_WAVE2_CONG_IGN. We use the shorthand W16 for WT_WAVE2_CONG_IGN in the formulas below.

1. Calculate baseweights for all congregations nominated by GSS 2006 respondents.

   The 2006 GSS design included subsampling of households, so not all households have the same weight. Therefore, we use the minimum of the weights of the nominating households in the numerator. For example if a congregation is nominated by both a subsampled household and a non-subsampled household, this weight counts the congregation as nominated by the non-subsampled household.

$$W16_0^{D,E,F,G} = \frac{\min_{i \subset c}(W_i^{2006})}{S_c}$$

2. Rescale weights.

   Many data analysis programs assume that the sum of the weights is equal to the sample size. Thus it is good practice to rescale the weights to the total number of cases, to ensure correct calculation of standard errors and confidence intervals. Without changing the relative weights between the cases, we rescaled the weights for sets D, E, F, and G so that the sum of the weights is equal to the number of cases.

$$\beta = \frac{|D| + |E| + |F| + |G|}{\sum W16_0^D + \sum W16_0^E + \sum W16_0^F + \sum W16_0^G}$$

$$W16_{final}^{D,E,F,G} = \beta * W16_0^{D,E,F,G}$$

WT_WAVE2_CONG_DUP: Weight for 1,254 cases nominated in the 2006 GSS (SET=D, E, F, G), taking account of duplicate nominations. This weight allows users to analyze the data at the congregation level.

WT_WAVE4_CONG_DUP will be used by researchers who wish to analyze the 2006 nominated congregations only. Note that only cases in sets D, E, F, and G have non-missing values of WT_WAVE4_CONG_DUP. We use the shorthand W17 for WT_WAVE4_CONG_DUP in the formulas below.

3. Calculate baseweights for all congregations nominated by GSS 2018 respondents:

   The 2006 GSS design included subsampling of households, so the sum of the weights of the nominating respondents cannot be ignored.

$$W17_0^{D,E,F,G} = \frac{\sum_{i \subset c} W_i^{2006}}{S_c}$$

   The numerator sums the weights of all GSS respondents who nominated a congregation.

4. Rescale weights.

   Many data analysis programs assume that the sum of the weights is equal to the sample size. Thus it is good practice to rescale the weights to the total number of cases, to ensure correct calculation of standard errors and confidence intervals. Without changing the relative weights between the cases, we rescaled the weights for sets D, E, F, and G so that the sum of the weights is equal to the number of cases.

$$\beta = \frac{|D| + |E| + |F| + |G|}{\sum W17_0^D + \sum W17_0^E + \sum W17_0^F + \sum W17_0^G}$$

$$W17_{final}^{D,E,F,G} = \beta * W17_0^{D,E,F,G}$$

WT_WAVE2_ATTENDEE: Weight for 1,254 cases nominated in the 2006 GSS (SET=D, E, F, G), allowing users to analyze the data at the attendee level.

WT_WAVE2_ ATTENDEE will be used by researchers who wish to analyze the 2006 nominated congregations only. Note that only cases in sets D, E, F, and G have non-missing values of WT_WAVE2_ATTENDEE. We use the shorthand W18 for WT_WAVE2_ATTENDEE in the formulas below.

3. Calculate baseweights for all congregations nominated by GSS 2006 respondents:

   The 2006 GSS design included subsampling of households, so the sum of the weights of the nominating respondents cannot be ignored.

$$W18_0^{D,E,F,G} = \sum_{i \subset c} W_i^{2006}$$

   The attendee weight sums the weights of all GSS respondents who nominated a congregation.

4. Rescale weights.

   Many data analysis programs assume that the sum of the weights is equal to the sample size. Thus it is good practice to rescale the weights to the total number of cases, to ensure correct calculation of standard errors and confidence intervals. Without changing the relative weights between the cases, we rescaled the weights for sets D, E, F, and G so that the sum of the weights is equal to the number of cases.

$$\beta = \frac{|D| + |E| + |F| + |G|}{\sum W18_0^D + \sum W18_0^E + \sum W18_0^F + \sum W18_0^G}$$

$$W18_{final}^{D,E,F,G} = \beta * W18_0^{D,E,F,G}$$

WT_PANEL12_CONG_IGN: Weight for 262 panel cases (SET =A, B, C, F, G), ignoring duplicate nominations, which allows users to analyze the data at the congregation-level.

WT_PANEL12_CONG_IGN will be used by researchers who wish to analyze the congregations nominated and interviewed in 1998 and 2006. Note that only cases in sets

A, B, C, F, and G can have non-missing values of WT_PANEL12_CONG_IGN. We use the shorthand W19 for WT_PANEL12_CONG_IGN in the formulas below.

1. Merge NCS-I completed cases with NCS-II completed cases.

2. Keep only those NCS-I respondents who also completed the interview in NCS-II (some may have been nominated again in 2006). These cases are in SET = C, F, G for 2006.

3. Calculate baseweights for all congregations nominated by GSS 1998 respondents (IGNORING nominations from 2006; this uses NCS-I data).

   The 1998 GSS design is an equal probability sample of households, so all households have the same weight. To be consistent with the newer weights, we keep the definitions the same, using the minimum of the weights of the nominating households in the numerator.

$$W19_0^{C,F,G} = \frac{\min_{i \subset c}(W_i^{1998})}{S_c}$$

4. Rescale weights.

   We scale the weights here so that the sum of the weights is equal to the number of cases.

$$\beta = \frac{|C| + |F| + |G|}{\sum W19_0^C + \sum W19_0^F + \sum W19_0^G}$$

$$W19_{final}^{C,F,G} = \beta * W19_0^{C,F,G}$$

NOTE: This methodology is equivalent to starting with WT_ALL4_CONG_IGN for the panel cases and re-scaling to the sample size.

WT_PANEL12_CONG_DUP was NOT created, probably because necessary data from NCS-I was lacking during NCS-II

WT_PANEL12_ATTENDEE: Weight for 262 panel cases (SET =A, B, C, F, G), which allows users to analyze the data at the attendee-level.

WT_PANEL12_ATTENDEE will be used by researchers who wish to analyze the congregations nominated and interviewed in 1998 and 2006. Note that only cases in sets A, B, C, F, and G can have non-missing values of WT_PANEL12_ATTENDEE. We use the shorthand W20 for WT_PANEL12_ATTENDEE in the formulas below.

1. Merge NCS-I completed cases with NCS-II completed cases.

2. Keep only those NCS-I respondents who also completed the interview in NCS-II (some may have been nominated again in 2006). These cases are in SET = C, F, G for 2006.

3. Calculate baseweights for all congregations nominated by GSS 1998 respondents (IGNORING nominations from 2006; this uses NCS-I data).

   The 1998 GSS design is an equal probability sample of households, so all households have the same weight. To be consistent with the newer weights, we keep the definitions the same, using the minimum of the weights of the nominating households in the numerator.

$$W20_0^{C,F,G} = \frac{\min_{i \subset c}(W_i^{1998})}{S_c}$$

4. Rescale weights.

   We scale the weights here so that the sum of the weights is equal to the number of cases.

$$\beta = \frac{|C| + |F| + |G|}{\sum W20_0^C + \sum W20_0^F + \sum W20_0^G}$$

$$W20_{final}^{C,F,G} = \beta * W20_0^{C,F,G}$$

NOTE: This methodology is equivalent to starting with WT_ALL4_ATTENDEE for the panel cases and re-scaling to the sample size.

Appendix A: Imputation of NUMADLTS

The number of regularly attending adults, NUMADLTS, is an integral variable in the calculation of the NCS weights. These data are collected from congregations during the NCS interview. When this variable is missing we must impute it from available data.

The method of imputation depends on the data that are available for a given congregation. We used the first of the following methods that we could:

1. If both NUMTOTAL (the total number of congregation members) and NUMREGLR (the number of regularly attending members) are non-missing, we used regression imputation to estimate NUMADLTS. That is, for the cases where all three variables are non-missing, we estimated a regression equation that predicts the log of NUMADLTS from the logs of NUMTOTAL and NUMREGLR. Then, for cases where NUMADLTS is missing, we estimated it using the coefficients from the regression equation.

2. If only NUMTOTAL is non-missing, we used regression imputation with the log of this variable only.

3. If only NUMREGLR is non-missing, we used regression imputation with the log of this variable only.

4. If both NUMTOTAL and NUMREGLR are missing, we used the variable collected in the nominating round of the GSS. This variable, CONGNUM, is the GSS respondent's estimate of the number of regularly participating adults at his/her congregation. We know that there is some bias in this estimate and that the error is larger for larger congregations, so we again took a log when fitting the regression model and deriving the imputation parameter.

5. If none of the above methods were available, we used mean imputation to fill in the missing values of NUMADLTS.

The above procedure was followed for imputing NUMADLTS in Waves II-IV. A somewhat less sophisticated approach was used for imputing NUMADLTS in the Wave I data.

The variable IMPSIZE flags the cases with imputed values on NUMADLTS.

Appendix B: Derivation of Optimal Lambda Parameter

From: O'Muircheartaigh, Colm, and Steven Pedlow. 2002. "Combining Samples vs. Cumulating Cases: A Comparison of Two Weighting Strategies in NLSY97." *Proceedings of the American Statistical Association*, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association, pp. 2557-2562.

To maintain the characteristic that the weights from both samples together sum to the population size (rather than each sample independently), the CX weights were multiplied by λ ($0<\lambda<1$), and the SU weights were multiplied by $1-\lambda$ in producing estimators based on both samples together:

$$\hat{\theta} = \lambda\hat{\theta}_c + (1-\lambda)\hat{\theta}_s$$

in which $\hat{\theta}_c$ represents a statistic derived from the CX sample and $\hat{\theta}_s$ represents the corresponding statistic from the SU sample. Because the two samples are independent, the optimum λ for a weight of this form is proportional to the relative effective sample size in the CX sample:

$$\lambda = \frac{n_c / d_c}{n_c / d_c + n_s / d_s}$$

$$1-\lambda = \frac{n_s / d_s}{n_c / d_c + n_s / d_s}$$

in which n_c and n_s are the nominal sample sizes for the CX and SU samples and d_c and d_s represent the design effects for the estimators from each sample. It is inconvenient to use the design effects themselves, since they will vary from one variable to the next. Instead, a general factor was used (one plus the squared coefficient of variation of the weights within each sample), as was done for NLS79; this factor captures the impact of unequal weighting on the sample efficiency:

$$\hat{d}_c = 1 + \left[CV(W_i \in \text{CX})\right]^2$$

$$\hat{d}_s = 1 + \left[CV(W_i \in \text{SU})\right]^2$$