
THE NATIONAL CHILDREN'S STUDY 2014: AN ASSESSMENT

USER DOCUMENTATION FOR THE COST ANALYSIS WORKBOOK

As described in National Research Council (NRC)/ Institute of Medicine (IOM) (2014, p. B-1) the Panel on the Design of the National Children's Study (NCS) and Implications for the Generalizability of Results "was charged with evaluating the proposed methodologies for the National Children's Study (NCS) Main Study, including whether such methodologies are likely to produce scientifically sound results. In trying to carry out this part of its charge, the panel found that the relevant cost information was not available. Thus, in order to assess alternative approaches with key design elements, such as sampling frame and design, the recruitment and retention process, and broad aspects of the interview schedule and data collection, the panel commissioned two consultants – Lisa Schwartz of Mathematica Policy Research and Randall Olsen of Ohio State University – to construct a cost model of several design options." The results of the cost analysis are described in Chapter 5 of NRC/IOM (2014). The assumptions behind the cost model are described in Appendix B.

The consultants developed the Field Budget excel workbook to support the analysis by the panel. As such, the workbook was designed for internal purposes. While the workbook was not designed for use by external users, the panel felt there were benefits to making the workbook publicly available. First, the workbook further specifies the costs associated with each of the key data collection assumptions (described in Appendix B). Second, it allows users to change to change some key data collection parameters and immediately to see the cost implications of those changes. This User Documentation provides a roadmap to the workbook and highlights where users can make modifications. Users who wish to make modifications should download and save a local version.

Approach to Cost Analysis

The cost estimates that were produced for the panel were built from the bottom-up. That is, the consultants first delineated the tasks, subtasks and activities that comprise the overall data collection effort. Based on their experience, they created assumptions about the level of effort needed to complete each of these component activities. As noted in Chapter 5, these assumptions were shared with the NCS Program Office and they agreed substantially with the assumptions, or at least did not disagree. Our bottom-up approach to estimation is reflected in the individual worksheets. For example, the component costs of recruiting, telephone interviewing and in-person (face-to-face) interviewing are shown on the worksheets labeled "NCS Recruitment 250 Hospitals 5 Providers," "Recruitment Travel," "Max Phone – No Sibs," and "F2F with sibs", respectively.

Other worksheets were developed to help model the costs of the design options considered by the panel. For example, "Duncan Fertility Facts" includes the data needed to inform the size of the sample of siblings that were included in all eight cost models described in Chapter 5 of NRC/IOM (2014). This worksheet was originally designed by Greg Duncan. The "Parameters" tab delineates the component costs of interviewing and reflects assumptions about the number of cases that will be relatively easy to engage in the interview and the number of cases that will require a more intensive effort. The "Costs Year 1-7 What-If" sheet is the main calculation sheet.

Brief Description of Worksheets

Each tab in the workbook is an individual worksheet. Formulas are used to link information across worksheets. Here, we describe each worksheet beginning with the leftmost tab “Max Phone – No Sibs.”

Max Phone – No Sibs: This sheet shows the costs of telephone and field interviewing for a sample of births, assuming any sibs that might be present do not materially change the cost of the phone effort¹. It assumes that phone is the primary mode of data collection and that field staff are only used to secure cooperation. It includes interviewer labor and other direct costs such as equipment, travel and per diem. The cost components on this sheet are used to build up the estimated cost per complete for a telephone interview (cell B51) which flows through to cell G17 in the “Parameters” spreadsheet. These estimates are used to inform the cost models.

Parameters: This sheet shows the costs associated with interviewing. It specifies the number of hours needed for various trainings and provides additional information about what is included in equipment costs. Columns B and C in this spreadsheet appear to be identical. However, Column B provides the cost assumptions that are used in the calculations in other spreadsheets. Users can make changes in column B to see which assumptions are major cost drivers. Column C contains the cost assumptions from the baseline case. It is recommended that Column C should be left unchanged so the user can accurately revert to baseline assumptions as needed. Cells F12-G18 of the Parameters sheet is a summary table that shows the cost of in-person (from F2F with Sibs) and telephone interviews (from Max Phone-No sibs.).

F2F with Sibs: This sheet reflects the costs of face-to-face (F2F) interviewing assuming that the study collects data on siblings born within the recruitment period to women producing sampled births. It shows the costs associated with field interviewing once the sample of birth mothers has been selected and recruited. It includes cost-savings associated with including siblings in the sample. The inclusion of siblings reduces recruitment costs and increases interviewer time in the household. Cells B6 and C6 in the Parameters sheet indicates that in the baseline model 11.6% of the sample would be made up of siblings and this would generate economies in securing cooperation and travel.

Data on the “F2F with Sibs” sheet is entirely formula driven. Users should make changes to the assumptions on the “Parameters” tab and will be able to see the cost implications on this sheet as well as in the tab “Costs Year 1-7 What-If”.

Duncan Fertility Facts: This worksheet was prepared by the panel and shared with NCS to verify that we are estimating the number of expected sibling births in a consistent manner. The Appendix to this memo reproduces the documentation of the spreadsheet that was sent to NCS along with the spreadsheet. This worksheet provides data on fertility that were used to calculate the number of births per year, including sibling births. Information on this sheet is used for computing the number of “Bumper Babies²” (described in “Costs Year 1-7 What-If” below)

NCS Recruitment 250 Hospitals 5 Providers: This sheet shows a detailed set of activities and the associated costs needed to recruit 250 hospitals as the primary sampling units (PSU) and five health care providers per hospital as the secondary sampling units. It includes the

¹ Sample size is given in Parameters B4. The percent of easy cases is given in Parameters B34, and the percent of hard cases is given in Parameters B35. The remaining sampled children are assumed to be non-respondents, even after extensive follow-up efforts.

² Sibling births in a given year.

costs associated with recruiting entities (hospitals and providers) and women (births) into the study. It shows the difference in costs for hospital versus practice-level recruiting. It includes labor hours and other direct costs.

Recruitment Travel: This sheet accounts for all recruitment-related travel costs. Column C on this sheet is linked to NCS Recruitment tab described above. Users may make changes to the number of staff, hospitals or practices on the “NCS Recruitment” tab. These changes, which affect the number of recruitment-related trips, will be reflected on the recruitment travel tab.

Costs Year 1-7 What-If: This is the main calculation spreadsheet. After setting parameters described below, results are displayed in cells I175-I198. The results associated with the panel’s 8 models, (model 1 being the baseline case proposed by the NCS Program Office) are displayed cells K173-R199. The parameters needed to reproduce those columns are shown in the table below.

Users can assess the effects that changes to certain assumptions may have on costs by modifying cells on this worksheet. Editable cells are highlighted in yellow and are commented.

- Cell C5 is the number of women assumed enrolled and giving birth in each of the four years for the four groups of sites³. Recall from the report that we assumed the PSUs would be put into four groups with the start years for each group being in one of four consecutive start years to reduce the peak costs of the project. The data collection effort is concentrated in the first two perinatal years due to their importance.

- Cell C7 adjusts the fraction of women enrolled prenatally as opposed to at delivery.

- Cell C8 adjusts the fraction of prenatal cases that receive a second prenatal interview.

We recommend the minimum value here be set at 0.2 to reflect the NCS experience with roughly 20 percent attrition after the first prenatal interview. Setting this value higher allows for the project design to accommodate the collection of additional, important, prenatal data. NCS project staff pointed out some mothers may be enrolled and then given a prenatal interview sufficiently late in their pregnancy that a second prenatal interview cannot be scheduled. Hence we used 0.7 to account for both late recruitment and attrition.

- Cell C9 allows for the possibility the study cannot interview all mothers in the hospital and recover all the information and specimens desired. This cell allows the user to vary the fraction of cases where a short hospital stay or other factors combine to require a home visit to complete the data collection protocol at birth. As noted in the Appendix, NCS originally assumed such home visits would not be needed, but the consultants doubted this would be the case and allowed for the need for some information and specimens to be recovered in the home shortly after birth.

- Cell C10 allows the user to include (value=1) or drop (value = 0) the 6 month face to face interview. An intermediate value would allow for a cost estimate in the case that only a fraction of those interviews are completed.

- Cell D10 allows the user to include (value=1) or drop (value=0) the 9 month (telephone) interview. An intermediate value would allow for a cost estimate in the case that only a fraction of those interviews are completed.

- Cell G1 controls the number of women enrolled pre-conception and then tracked. Cell G2 is the ratio of the number of women who must be recruited and tracked pre-conception divided by the number of those women who give birth during the enrollment period. It is based

³ This number is the number of births.

on NCS experience. Cell G3 is the number of home visits made to each woman in the preconception sample. Cell G4 is the assumed number of “check-in” or tracking telephone interviews for each woman. These costs (cell G5) feed into overall recruitment costs in cell Z11.

This sheet shows the costs for the first seven years of the study and allows users to see how sensitive each of the different models (described in the report) are to various design assumptions. Scroll down to see summary costs by model by year and graphs of each model’s costs over time. The columns with the words “Bumper-Baby” are used to project the number of siblings born during the recruitment period to mothers who were first recruited at a provider or at delivery.

The fields in blue provide particular numbers of mothers enrolled per year to generate particular numbers of children over all four recruitment cohorts. Note that the birth and enrollment of subsequent siblings increases the number enrolled beyond the number of mothers enrolled. The fields in yellow (C5 – C10 except C6, D10, and G1) drive the remainder of the projections. Toward the bottom of the worksheet (beginning around row 175), fields in yellow allow one to alter assumptions about sample retention as well as cost of living adjustments (COLA). Changing fields that are not in yellow will change the model in unpredictable ways.

Table 1 illustrates the cell values used by the panel to generate the cost profiles reported in NRC/IOM (2014).

We only varied the cells in yellow to generate the various scenarios in the report, so users who alter any other cell in the workbook should not be surprised that they cannot replicate our cost estimates.

Table 1. Parameters Used to Produce Model Results

Model	Values for Cells to Reproduce Model Results						Value of C5 for Cost neutrality with 1
	C5	C7	C8	C10	D10	G1	
1	5525	.50	.7	1	1	5000	NA
2	5525	.97	.7	1	1	5000	4928
3	5525	.97	.7	1	1	0	5319
4	5525	.97	.7	0	0	0	6295
5	5525	.97	.7	1	0	0	5539
6	5525	.97	.7	0	1	0	6013
7	5525	.97	.2	1	1	0	5558
8	5319	.97	.7	1	1	0	5319

Reference

National Research Council (NRC) and Institute of Medicine (IOM) (2014). *The National Children’s Study 2014: An Assessment*, Panel on the Design of the National Children’s Study and Implications for the Generalizability of Results, G.J. Duncan, N.J. Kirkendall, and C.F. Citro (eds). Committee on National Statistics, Division of Behavioral and Social Sciences and Education, and Board on Children, Youth and Families, Institute of Medicine. Washington, DC: The National Academies Press.

NICHD 2013d

Documentation for Worksheet ‘Duncan Fertility’

NRC/IOM panel

Number of sibling births in the NCS

Background

A key parameter for the NCS is the number of siblings that would be included in the planned 90,000-child main study. NCS’s estimate is 8,000. The panel prepared this worksheet to estimate this number independently. This spreadsheet estimates that there would be 7,935 non-twin births during a four year recruitment period. The spreadsheet also provides an estimate for the number of sibling births during a longer 7-year recruitment window, but that is not discussed in the NRC/IOM(2014).

To understand our logic, it is helpful to break down the calculation for the initially-enrolled babies who are first births and initially-enrolled babies who are second and higher-order births. Vital statistics data (http://www.cdc.gov/nchs/data/nvsr/nvsr62/nvsr62_09.pdf, Table 2) show that 39.7% of all births are first births, so we employ the assumption that 39.7% of target births in the study will be first births and the rest will be higher parity births. We perform separate sibling calculation for each group and take a weighted average of these two sets of numbers.

Siblings born subsequently to first births during the four-year recruitment period:

Let’s begin with the first births and detail our calculations of the likely number of siblings born over four and seven year periods. We concentrate most of our comments on the four-year calculations.

First study year: We assume that the average birth in the first year of the study occurs six months into the first year’s study period. The only chance of a sibling for that birth in the remaining six months of the first study year is with twins, triplets, etc. The birth rate for twins was 3.3% in 2012 (http://www.cdc.gov/nchs/data/nvsr/nvsr62/nvsr62_09.pdf) and for other multiple births is around .2%, so in the first study year you get 3.5% twin+ siblings. This figure is entered into cell B3 of the attached table.

Second study year: From Table 7 of NICHD (2013d, p 21) (which comes from the “Total, 2006-2010” row of Table 9, p. 22 of <http://www.cdc.gov/nchs/data/nhsr/nhsr051.pdf> and is reproduced in cells A30-B34 in the attached spreadsheet), it appears that in the second year of the study you would get about half at the 2.5% “12 months or less” rate and half at the 13-24 months’ 17.5% rate, which averages 10.0% siblings born in the second study year – a figure that is entered in cells B4 and B31.

Third study year: In the third year of the study you get the other half of the 13-24 months’ 17.5% rate and half of 25-36 months’ 17.2% rate, which averages 17.4% siblings born in the third study year, which is entered in cells B5 and B32.

Fourth study year: In the fourth year of the study you get other half of 17.2% and half of 12.3%, which averages 14.8% siblings born in the fourth study year, which is entered in cells B6 and B33.

Since we are concentrating on a four-year recruitment window, you can ignore the following:

Fifth study year: In the fifth year of the study you get other half of 12.3% and half of an assumed 9.0% rate which averages 10.7% siblings born in the fourth study year, which is entered in cells B7 and E34.

Sixth study year: In the six year of the study you get other half of the assume 9.0% rate and half of an assumed 6.0% rate which averages 7.5% siblings born in the fourth study year, which is entered in cells B8 and E35.

Seventh study year: In the seventh year of the study you get other half of the assume 9.0% rate and half of an assumed 6.0% rate which averages 7.5% siblings born in the fourth study year, which is entered in cells B8 and E35.

Target children born in the second year of the four-year recruitment window will have a similar pattern of subsequent siblings, except all of the numbers in rows 3-9 are displaced downward by one row. Similarly for sibling births to target children born in the third and fourth year of the recruitment windows.

The “Mean” column (F) takes a simple mean of the preceding four columns. The F11 cell sums the fractions in the first four years to get the estimate that **23.36% of first births occurring the four-year recruitment interval can be expected to be followed by a subsequent birth within that same four-year interval.**

Siblings born subsequently to first births during the four-year recruitment period:

So far so good for first births. But only 39.7% of the births in the first study year will be first births -- the rest will be higher parity births with considerably fewer subsequent births. We couldn't find counterpart NCHS data on birth spacing for 2nd-3rd births, 3rd-4th births, etc. although we saw a scrap of Google books information from *The Population of the United States: 3rd Edition* By Richard E. Barrett, Donald J. Bogue, Douglas L. Anderton indicating that a .5 fraction applied to first births would be a reasonable assumption. Please let us know if you have found a more recent and reliable source.

The entries in cells H3-L9 and L11-L12 simply multiply the .5 fraction times the entries in cells B3-F9 and F11 and F12. Based on our 50% assumption, the entry in L11 shows that 11.68% of 2+ parity births are expected to be followed by a subsequent birth within the four-year interval. Thus, **11.68% of higher-parity births occurring the four-year recruitment interval can be expected to be followed by a subsequent birth within that same four-year interval.**

Bottom lines

Cell L15 takes a weighted average of these two first and higher-parity to set our estimate that **16.32% of all births occurring the four-year recruitment interval can be expected to be followed by a subsequent birth within that same four-year interval.**

Ignoring attrition and assuming a 90,000 total sample size and 16.32% of target births are followed by sibling births, then **the 90,000 will be composed of 73,374 (90,000/1.1632) target births and 12,626 (the remainder) sibling births.**

Applying a 20% eventual attrition rate reduces these numbers to 61,899 and 10,101, respectively. Applying the 3.5% rate of twin births to the 10,101 gives an **expected number of twin+ sibling births of 2,166 and 7,935 non-twin births**, which hits the 8,000 number pretty much on the nose.