

Quality Assurance

Quality assurance (QA) saves time and improves data quality, whether conducting fieldwork or cleaning and analyzing the data. Make sure to stress it when training and when conducting a survey, as it should be a part of every step!

Below are some recommendations on how QA is done at each step of the process, with some examples. However, this is by no means an exhaustive list, so build in whatever structures and quality checks you feel are needed. We recommend tailoring the QA protocol to your group as some groups may need more or less supervision and quality control checks than others.

Enumeration

- Make sure that the method you use to classify all of the food outlets is used consistently by everyone (e.g., if a green highlighter is to denote all fast food restaurants, make sure that everyone uses a green one instead of a different color), so information can be correctly entered into a spreadsheet.
- Make sure that all of the “don’t know” classifications are followed up on either by internet searches or phone calls.
- Once all of the food outlets are entered into a spreadsheet, review to make sure no information is missing (e.g., address or phone number).
- Make sure to have agreed upon procedures explained to the field raters before beginning fieldwork, on how to handle any discrepancies with the enumeration list (e.g., when to get a new id number).

Fieldwork

- You may want to go out with your raters to rate a store and a restaurant together for practice. When you return to the office, review their survey forms – be sure that:
 - All items are completed
 - All items are correctly completed by checking their survey responses against yours
 - The handwriting is legible and their notes/comments make sense (you don’t want to have to ask someone to interpret what they wrote about an item in a store, 40 stores later!)

Provide feedback to them on survey completion and the criteria for some of the measures, if needed.

- Stress to raters that if they have any questions out in the field, to note them down, so they can be asked once they return to the office. Most likely, if one rater has a question about a particular measure or food item, others do as well. **Make sure to share the questions and answers with everyone at your staff meeting or by email.**
- Also, you may want to perform reliability checks on the data by having a certain percentage (5-10%) of stores and restaurants surveyed by two raters to see if they have similar responses. Comparisons of inter-rater agreement can be done
 1. by manual inspection/review immediately after fieldwork,
 2. by calculating % agreement and kappa statistics after creating a database, or
 3. both ways.

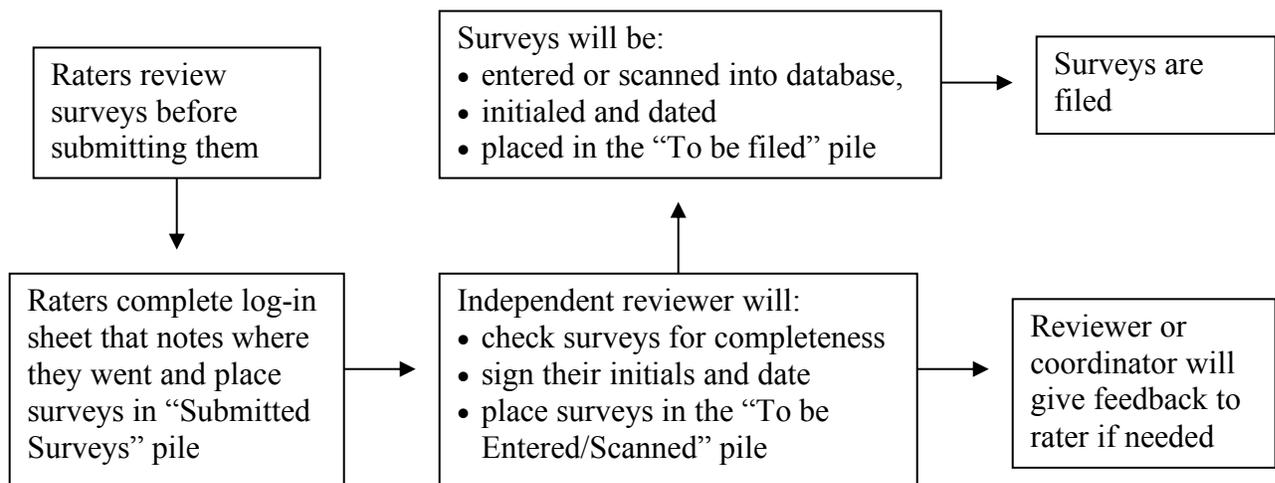
Quality Assurance

- Make time to check the survey forms **routinely**, especially in the beginning, so any survey form errors that may occur will be caught before too many have been completed.
- Have clear procedures for all aspects of the survey: from where everything is located for going out to do the fieldwork to the procedures for when they return from the field with their surveys.
- **Make sure raters leave enough time to thoroughly review each survey they complete during an outing.** This may mean not completing their field assignment list for the day. Stress that you'd prefer this to having only partially reviewed or rushed surveys.

****Go by the assumption that a rater may not return the next day to the office, so make sure that everything is filled in before s/he leaves for the day.****

Data Management

- Make sure that the flow of the data is clearly understood from the time the completed surveys enter the “to be reviewed” pile, to the “data entry” pile, to the “entered data” pile. You may want to develop a tracking spreadsheet or paper coversheets to make sure that each step in the data management process has occurred. Do as much or as little as you need to in order to make sure that the all of the surveys have been completed, reviewed, and entered or scanned.
- You may want to keep the surveys **categorized** by ID number, area or whatever makes sense to you, in case you need to go back and locate one to verify or correct an answer in the database.
- Below is an outline of the **basic process**:



- As mentioned above, performing a few reliability checks can circumvent many headaches! In addition to performing reliability checks between/among hardcopy rating forms, it's a good idea to also **perform periodic reliability checks between hardcopy and data in the database/spreadsheet.**

Quality Assurance

- Depending on how many restaurants/stores and how long the data collection/entry process takes, you may find it advantageous to have a **Database Log** in which you record any coding schema changes or other revisions/corrections to the database information. For example, with the previous NEMS data coding scheme changed ~ first part of study a blank was left blank for a particular question; second part of study that blank was coded as a 0 ----- so, at the end of the study there were blanks and zeros in the database, but conceptually they meant the same thing. It might be best to keep a systematic record of such changes in coding schemes, and additionally, it might have been best to retroactively change all the blanks to zeros.
- If there are changes/corrections/etc. made to a database, it is not a good practice to overwrite the original file. Stated differently, **after a database is revised, save it under a different name than the original file**. This way no data are ever lost and the evolution of the dataset can be traced (which is especially helpful when there are issues or questions like “where did this number come from?”, “why is this store coded this way but this store is coded that way?” . . . and many, many more!)
- Related to the comment above, we suggest adding the date to all file names. Example: NEMS DATABASE 03.02.2006. Then, each time the database is revised in any way, the new database is saved under the current date. This produces a clear data trail, and there will be no question which file is the most-current.

Data Analysis

- Have a clear **plan for analyses** ~ list of empirical questions to be investigated to guide analyses is a good place to start.
- Make sure that the statistician and the project staff are always on the same page.
- Have a very clear **Data Dictionary** that links the questions and responses on the hardcopy to the numeric/string codes in the database. If the statistician creates any variables (computations, etc.) and saves them in the database, add that information to the Data Dictionary.
- Have the statistician keep one, **updatable log** of each manipulation done - - - empirical questions being investigated for those particular analyses, analyses used, results of analyses (statistical), interpretation (in real words), file/path of database used, file/path of program file (if applicable), file/path of output file, date.
- The first thing in that plan should be data screening and cleaning. Despite all of the checks and balances, it is still possible that information has been miscoded or reported incorrectly. Here are a few things you can do to identify data issues:
 - **Run minimum/maximums for all appropriate variables**. Make sure that these min/max are possible given your coding schemes. For example, the number of healthy salads at a restaurant cannot be a negative number, so if the minimum is negative for that column of data there is a problem. If the number of tables in the restaurant is a huge number, then you should probably recheck that column . . . lest one of the restaurants is the size of the Superdome.

Quality Assurance

- **Do simple frequencies** - - - number of restaurants, sit-down, take-out, this neighborhood, that neighborhood, etc - - - - Check these frequencies with those reported in the Field Log. If you had multiple raters for some stores or restaurants, you should randomly select one observation per store for this analysis.
 - **Check to make sure that your frequencies make sense in relation to other frequencies** . . . more specifically, a restaurant can't have more healthy salads than it has salads. There are a number of places where this type of check should be performed. (Done easily using SPSS or .xls. For this example, run a computation dividing what should be the larger number by what should be the smaller number - - - if the resultant number is > 1 , then there is a problem.)
- During analyses, make sure that the programs used by the statistician do **not** overwrite the data file. For example, if a SAS program is used to recode a sub-set of data, the resultant database should be saved under a different name (or at least the current date) from the original data file.