

## OHD Outbreak Investigation Templates for Questionnaires, Data Entry, and Data Analysis

### Overview

This describes how to use a family of questionnaire templates for outbreak investigations. Although most commonly used with outbreaks of gastroenteritis, they can be adapted for other kinds of investigations. Although they can be used as stand-alones, they are optimized for use with a FileMaker data entry template, "Data Entry.FP5". (Note to non-FileMaker users: although a tedious project, these templates could be replicated in Access or most other database programs; help yourself.) The FileMaker template allows for fast and accurate data entry, and feeds directly into pre-written EpiInfo analysis programs. Because the set-up does not vary much from outbreak to outbreak, data entry people quickly become (and stay) proficient.

The key difference between this system and more traditional data entry is that many exposures (typically food items), as well as symptoms and other characteristics, are grouped into blocks for data entry. Only "yes" answers are entered; "no" is otherwise assumed. For example, if there were 20 items available at a suspect meal, traditionally one would make 20 yes/no variables. In this system, however, those 20 items are entered as a single variable. If the person ate only items A, C, and H on the list, for example, one would enter "ACH".

Behind the scenes, these raw data are being split into "traditional" YES/NO variables. At any time, with the flick of a button, the data can be automatically cleaned and exported (e.g., as DBF files) for analysis in EpiInfo or other programs. A companion EpiInfo 6 program (*program.doc*) gives you symptom profiles, epi curve details, and 2x2 tables for each food item – all you need to do is specify the file names and delete unused variables. Generic variable names are used [e.g., item\_1a, item\_1b, item\_1c, ...] so that the same program can be recycled with little or no modification.

Data cleaning can be done at any stage: in FileMaker, in Excel, in EpiInfo or SPSS – as you prefer. Some cleaning (e.g., standardization of onset and sx duration times) is done semi-automatically; there are many checks built into the data entry FileMaker file. Manual overrides are available at all steps. Data entry queries (e.g., illegible or inconsistent answers) can be electronically flagged for later review. Other features are available to the more advanced FileMaker user.

### STEP 1 Customize the Questionnaire in a Word Processor

Work on a copy of the template. The template was created in MS Word, and it helps a great deal if you know how to control formatting in that program. There is a lot of embedded formatting in the template (e.g., line spacing, tabs, page breaks, paragraph borders), and if you are careless you can end up with a real mess on your hands. The typefaces (fonts) used are Palatino, Arial, and Wingdings (the latter for the boxes). You're on your own if you work on this in some other word processor. It should work, but you'll need to be very alert to formatting subtleties.

You'll almost always want to collect certain core variables--name, age, and sex, but there are also some placeholder (dummy) demographic variables that you can use as necessary – 6 text and 4 numeric: text1, text2, text3, text4, text5, text6, numeric1, numeric2, numeric3. The variable called "match" can be used if this is a matched analysis--enter the patient's name or id number that you are matching to. Examples of variables to consider include school or class name, grade, Scout troop number, etc. The text in red is stuff that you will almost always want to customize; the color is just a reminder. When you're done, you will presumably want to turn everything black! (Select all and choose black text color.)

No, we didn't forget to include a space for the person's NAME. It is tucked away in a page header, so it will print at the top of every page – a nice insurance policy in case pages get separated.

Use the blocks to organize and list food and other exposure items. Blocks can be used for individual meals, or to split up meals with many food choices (e.g., large buffets) into logical groups (e.g., desserts, salad items, meat entrees, pasta, etc.)--whatever makes sense. You can use up to 16 blocks without modifying the FileMaker templates, but many questionnaires only need one or two. The little letters are keypunch codes; they are necessary for rapid data entry. *Delete them at your own risk!* One can use up to 25 items per block (letters A-Y). If you keep the blocks relatively even in length across a row, it minimizes wasted space, but don't let aesthetics get in the way of good organization in terms of asking the questions. At the same time, remember that in general the fewer the number of blocks, the faster the data entry. The blocks are shown going from top to bottom, left to right, but there is no reason why you couldn't put one data entry block across two or three columns. The only real limit is no more than 25 items/block.

Here are some *examples* of how to organize the blocks. They may give you ideas about how to use them.

- **Weekend Camp** (church, Girl Scouts, etc.) Typically multiple meals with not a lot of variety at any one meal. Blocks could be: Friday dinner, Saturday breakfast, Saturday lunch, Saturday snacks, etc., or, if there aren't many items at each meal, a block could be all the meals for one day.
- **ALC/jail/institutional outbreak.** Much the same as the weekend camp scenario, really. Ask about enough meals to blanket the likely exposure period: one meal, one block. Remember that if a case eats everything in a data entry group (which is common at jails), you can just keypunch a shortcut (e.g., "Z") and FileMaker can translate that into all of the individual letters for you (or whatever set of letters you define). Of course, these data entry aids must be explained to the data entry people.
- **restaurant outbreak.** Follow the menu: Appetizers, Salads, Pasta, Entrees, Desserts, Beverages. Tip: for many big chain restaurants and fast-food places, there may be information on their national web pages about what's on the menu. (Or get someone with a cell phone to stand in the restaurant and read it to you off the board.) Again, you can combine multiple (short) lists into a single block. Use subtitles to separate the subgroups, and insert an extra carriage return in the checkbox column.
- **Potluck party.** Organize by food group, or perhaps by location (the table at the north end of the cafeteria, the one at the south end). A common problem is that there may have been 4 kinds of fruit salad and 3 lasagnes, but no one can remember which was which. Do the best you can (e.g., the lasagne in the red dish), but consider adding an extra variable like "any fruit salad" or "any lasagne" in addition to the specific variables (Marge's fruit salad, the fruit salad that had marshmallows in it, etc.). These variables will be watered down (less statistical power because of misclassification) but may still prove useful--in any event better than nothing (which may be your only alternative).
- **Wedding reception/dinner party.** Typically only 1 meal, which usually fits into a single block.
- **Shotgun questionnaires.** *Salmonella* Heidelberg is up all over, for example; what is the common source? A separate questionnaire template (shotgun.doc) is available for these situations. (Such a questionnaire would almost always be created by ACDP, often in conjunction with other states.)

This template includes examples of exposure grids that are arranged 1, 2, and 3 blocks across. If you need more rows, just copy and paste. Choose whichever works best for you, or mix and match--but keep in mind that a compact layout is a plus. Don't delete the little block i.d. numbers. Delete unused items and blocks.

The existing FileMaker data entry template allows for up to 16 exposure blocks. At 25 variables, that's 400 yes/no items, which should be plenty. For most outbreaks you won't need more than 10-30 items in 1-4 blocks. *Almost all outbreak questionnaires should fit onto 2 or 3 pages--1 or 2 for exposures and another with the clinical information for the sick people only.* In other words, most of the template typically gets deleted. Keep in mind that some programs (Excel, SPSS?) limit the number of columns

(fields/variables) that they can import – typically to 256. Variables in excess of the limit are cut off. EpiInfo 6 does not have this limit.

Whenever possible, I like to reserve the last page of the questionnaire for use with sick people only. So the ultimate question on the penultimate page is usually “Have you been sick at all since ....?” For those who stayed well, that is the end of the questionnaire. That way, you never have to use/fax the last page for controls.

Individual questions about meal time are obviously unnecessary if everyone ate at roughly the same time. If there are multiple meals, this is more complicated, and you’ll have to improvise. There are spaces for up to five meal times, but the analysis program doesn’t automatically do anything with them. You’ll have to modify your analysis accordingly if you want to use them.

If you want to have a caller introductory spiel, write it out on a separate piece of paper. Don’t copy it onto every questionnaire--it just wastes space. Consider pasting a little calendar onto the spiel page (or the questionnaire, if it fits) as a memory jogger for onset dates. Sample calendars are provided.

Remember that this is a TEMPLATE. You are *supposed* to modify it. Keep in mind, however, that some of the material and much of the formatting is there for a reason. If you strike out on your own without understanding how the data entry and analysis templates work, you may be making your task much harder than you anticipated. If you add questions without keypunch codes, for example, you will make your data entry and analysis chores *much* more difficult. Use your imagination and the information that you have collected to make a questionnaire that serves the purposes you have as efficiently as possible. Don’t be in such a rush to make up a questionnaire that you leave out the key questions! If possible, pilot the questionnaire. (Ideally, this should be done with people that will be excluded from the analytical sample.)

Before you distribute the questionnaire, it is highly recommended that you save it as an Acrobat (PDF) file, specifying to “embed all fonts.” This is the ONLY way to ensure that the formatting, including page breaks, etc., doesn't get screwed up in e-mailing it around to people who may not have the same fonts and software that you do. If this is not possible (e.g., you don't know how to make PDF files, or they don't have e-mail), get someone to do it for you or – at a minimum – have the person fax back to you what they print out on their end *before* they start to use it.

## **STEP 2 Administer the Questionnaires**

Callers should be familiar with the questionnaire before they begin calling. If you have questions or don’t understand how to ask a particular question (or what it means), get it straight before you call. All interviewers should have a common understanding of how to interpret answers.

Use dark ink, and preferably nothing fine tipped. You want it to fax well.

If asked, it’s OK to answer reasonable questions about the outbreak, but avoid details (particularly about possible vehicles) until after you’ve gone through through the questionnaire. “I don’t want to influence your responses, so could we go through list of questions first, and then I’ll answer that, ok?”

Be sure to emphasize that you want to hear about foods that they just tasted, even if they didn’t eat but a bite or two. (For some people, that doesn’t constitute “eating.”) It’s fine to write notes in the margin if you have to.

Onset date and time. Nothing gives us more grief!! An estimate (if necessary) is OK, but try VERY HARD to get a specific date and a specific time out of them. If you ask “what time did you first start vomiting?” and they say “morning”--don’t leave it at that!! We can’t graph “morning” too well. Ask follow up questions as necessary: “About what time was that?” If they say “between 2 and 3 am” write “~2:30” in the AM slot. If they say “midnight,” make sure it is unambiguous which day you are talking about. The stroke of midnight after Friday evening is still Friday--but right after the stroke of midnight

it is Saturday. Don't write down what they say (necessarily)--write down what they mean--so keep asking until it is unambiguous.

For bugs with longer incubation periods, you might not care about the exact time of onset. If knowing the date is specific enough, you can delete the questions about time of onset to make things go faster.

### **STEP 3 Prepare for Data Entry**

If you use the FileMaker system, you'll need to modify the data entry layout by deleting unused fields and revising the labels on the layout. Make sure the tab order is correct; this step is often overlooked and really slows down data entry. (Hint: go to Layout mode and choose Tab Order from the menu....) You can change all the names on the screen, but don't change the underlying variable names in the Define Fields menu; the EpiInfo analysis program depend on the use of fixed variable names.

Of course, if you aren't going to use the FileMaker system, you can do whatever you like. The questionnaire template is just a word processing file; it can be used without FileMaker.

### **STEP 4 Data Entry in the FileMaker template ("Data Entry.fp5")**

Enter the data in FileMaker. Once you get the hang of it, it shouldn't take more than a minute or so to enter each record. More than one person can enter data at the same time. It works best if the file is located on the FileMaker server, but only Stephen and Bill can do that. Alternatively, anyone can open the file on their own machine (local hard drive or shared directory) and others should be able to see it when they choose the the Open/Hosts/ option. Yet another way to distribute data entry is for each operator to work independently on a copy of the database. When data entry is complete, merge the various files by exporting and importing. Make sure you "match by variable name" when you do the the import. If that doesn't mean anything to you, you should either read the manual or

### **STEP 5 Analyzing the Data**

If you use the FileMaker data entry template, variable names don't change from outbreak to outbreak, and one can use the standardized programs for analysis with only trivial modification. Analysis program templates have been developed for use in EpiInfo 6 (Bill), EpiInfo 2002 (Stephen), and SPSS (Barb); use whichever you prefer. The EpiInfo 6 program (at least) provides a complete demographic and clinical summary tailored for the outbreak summary report form, data for epi curves, as well as 2x2 tables for all exposure variables.

In EpiInfo 6, open the program file ("program.doc") in a word processor and make the following changes:

1. Write in the name of the data source files and output files (i.e., change "read o:/xxxxx.dbf to i:/outbreak/2003/2003-66.dbf – whatever).
2. Delete all the unused variables. The program creates 2x2 tables for all of the exposure variables – that's blocks 1-16 (A through Y each). Presumably you didn't use most of these variables, so why wade through a bunch of worthless junk?
3. (optional) If you don't like the generic names, you can recode them in EpiInfo. (This is probably only worth the time if you are going to be running this over and over again.) For example, if item\_2d is spinach, try this:

```
define spinach ___ (NB: make this at least 2 characters long.)
let spinach = item_2d
```

You'll have to make some other changes if you are doing stratified or matched analyses, or if you want to play around with any of the customized variables (e.g., text1, number2), or if you want to calculate incubations based on some time other than "meal 1". Use the search-and-replace function in a word processor to make big changes. For example, to stratify by the agegroup variable, replace "itemxxxx

status" with "itemxxxx status agegroup". (Actually, you would be replacing "status" with "status agegroup" in the word processor. This is very easy – if you know how to use the word processor well. Keep in mind that if the exposure time is uncertain, the incubation time is undefined, regardless of what the printout says.

#### **STEP 6      Clean it Up**

Take a couple minutes to clean up the Epi Info output before you print it out. It is just a word processing file, so you can open it in MS Word or whatever and delete all the garbage you don't care about. Add page numbers and the time/version of the analysis. Add labels (like the real food names) to the 2x2 tables of most interest, etc. Make the columns line up. For best results, always print in a monospaced font (e.g., Courier). Otherwise, things won't line up properly. Consider printing 2 or even 4 pages on a sheet.

#### **STEP 7      Think about It**

2x2 tables alone rarely are the sole solution to an epidemiological mystery. Ask yourself if the data make sense. Consider the effects of bias and confounding, and think about other ways to slice the data.

#### **STEP 8      Getting More Information**

For help using this system, contact us by phone (503/731-4024) or e-mail. Not everyone at the Health Division is expert at this system; you might want to start with Bill Keene, June Bancroft, Barb Progulske, or Beletshachew Shiferaw. Interesting, all have names starting with a "B". Obviously you should read these instructions first....

THE END