Working through the Data

STEP 6

Step 5 provided ideas to improve the quality and quantity of data gathered for evaluation purposes, including use of standardized and non-standardized instruments, qualitative and quantitative techniques.

Step 6 talks about what to do with the raw data after it has been collected, how to deal with it physically as well as how to uncover patterns related to program impact by simple statistical or qualitative analysis. Many community programmers are knowledgeable about methods of handling and analyzing data. The basic information and examples in this section will help provide instruction for junior or inexperienced staff and volunteers.

6.1 Drawing conclusions about program effects

In evaluation studies, a greater range of questions can be asked and answered when information is collected about individuals, then analyzed by grouping responses into categories.

In community programs, it is becoming common to compare data about individuals from before and after the program. This can provide persuasive evidence about the quality and likely effects of that program on participants, but cannot prove that the intervention caused the effects. Some research into longer-term interventions has produced evidence suggesting a causal relationship because the effect increases with the extent of participation, similar to what physicians would call a 'dose-response' relationship. Longer programs that can collect data about the same individuals regularly over time have more likelihood of showing such a relationship.

Even without making a claim for cause-and-effect, community programmers can learn much about their programs by looking at the extent of change across individuals, comparing amounts of change and looking at the full range of results to find frequency patterns, how often certain results are obtained, when and by whom.

6.2 Protect and preserve the data

Primary rules

- keep the raw data, composed of notes, logs, checklists, questionnaires, interview reports, etc. intact and secure
- make sure the secure place is recorded and known to more than one staff member, with safeguards are in place to monitor access to the original data
- produce an identification number for each original data collection form to make it easier to identify data, while preserving the informant's anonymity
- work from copies of the original documents
- when scanning or entering data to a computer, keep the original copy intact so that it will never be changed or overwritten
- identify each working file to keep track of progress, corrections etc. and make backups

Types of identification

- range from a simple sequential list of numbers (for computer entry add zeros to allow sufficient digits to accommodate the largest number)
- to serial numbers identifying the particular form (focus group, pre-test, post-test questionnaire etc.)
- numbers rather than combinations of letters and numbers are preferable for computer analysis

6.3 Review data for errors

Sources of error

Quality of data is improved if errors can be corrected at the source, e.g. when self-completed questionnaires are turned in during a program session.

Review forms immediately for obvious errors e.g.,

- missing data
- several boxes checked instead of one
- illegible writing

Ask the respondent to correct the problem.

Other sources of error

- inaccurate transcription of data
- misreading replies and
- keying errors when coding data into a data base or spread sheet for analysis

The best strategy is to try to solve problems in the data before tabulating and analyzing responses.

6.4 Organize qualitative data in relation to evaluation questions

Asking a few important questions of the data, and doing that analysis carefully, is the best use of time and resources for evaluation

The prime motivation for data collection and analysis is to learn more about what is going on in programs so they can be improved. As in other elements of community programming, it is important to set priorities for what can be accomplished within budget and resource limits.

After reading through at least some of the qualitative data collected, start to think about ways of relating the data to evaluation questions and indicators.

Become familiar with the range of responses and ideas raised. Data will raise questions that seem important to try to analyze, ideas that may be generalizable across respondents, (extent to which responses reflect the whole group or a specific subgroup.) By organizing and coding qualitative data, a researcher can compare responses among sub-groups (e.g., by gender or other known demographic data) or among responses and variables.

- Looking first for information needed to answer questions raised in the planning stage:
    - increases the speed of handling data and interpreting results
    - avoids the risk of being overwhelmed by reams of data, charts and tables

- Develop priorities for which questions are most important to the original intent of the evaluation.

The disadvantage of a narrow focus is that you can fail to notice unexpected outcomes.

Questions to ask while reviewing the data:

1. What information do the responses give about what the evaluation was assessing?
2. What ideas are common to responses?
3. How can responses be grouped to say something meaningful about the program—meaningful to people in the program, people running the program and the broader community?
4. Do responses say anything surprising?
5. Do responses from one question support or contradict those from other questions and from other data sources data.

Working through qualitative data: An example

A community recreation centre offering a recreation/homework program might ask parents an open-ended question: "Why did you enroll your child in this program?" and obtain the following responses:

a) to give him something to do instead of video games at home
b) to keep her out of mischief till I get home from work
c) to relieve my mother who looks after the baby
d) to learn English more
e) access to a computer for schoolwork
f) for homework help
g) to make sure he does his homework before dinner
h) to get more physical exercise
i) the teacher suggested it to help with homework
j) to make friends
k) he thought it would be fun instead of staying home
l) for the sports activities
m) an opportunity to socialize with other children
n) to get her away from the television

First, think about the reasons for asking this question. Perhaps programmers developed the program to keep latch-key children off the streets while their parents worked. They advertise the program as providing a positive alternative to hanging out on the street. Only response (b) specifically supports the contention that the program

helps latch-key children. Two other responses (c) and (g) might be categorized as 'parents wanting adult supervision for child,' which is similar, but clearly those children would not be out on the street without the program.

Programmers want to assess the fit between what they offer and parents' needs in order to better serve the community. They also wish to find out what elements attract participants so they can emphasize those in advertising and presentations to increase their numbers.

With those questions in mind, programmers decide how best to categorize the data to focus the analysis.

> A category like 'seeks productive activity for child' would fit almost all the responses but tells little about how to improve the fit with parents' needs or attract more children. Looking at the responses with those objectives in mind, one might group and code the responses as 1. "to provide academic help" (d,e,f,i), 2. "to provide supervision" (b,c,g), 3. "for outgoing social activity" (j,k,m,), 4. "to increase physical activity" (h,l). Responses (a) and (n) do not provide enough information to put them in either of the last two categories though one might interpret them as fitting in one or the other. They could go into a catch-all category 5. "other", or "alternative to passive activity at home."

The level of interpretation to be used with the data is set by the quality and amount of data collected, as well as original discussions with stakeholders about what they want to learn from the evaluation.

An alternative, using less interpretation, is to collapse fewer responses, giving each a separate coding number except for those with almost identical wording. That would mean collapsing only (f) and (i), 'homework help' and (j) and (m) 'to make friends.'

6.5 Coding and data entry

Making sense out of data entails counting responses and tabulating how frequently each occurs, which requires data to be coded and organized.

In evaluation studies, information is collected about individuals, then analyzed by group to find patterns and trends.

Coding

- code individual responses in accordance with a master coding sheet to allow easy computer identification, e.g., male = 1, female = 2; no = 0, yes =1.
- responses already in the form of numbers, as for age or grade or numeric scales (1-5 from least… to most…), are entered as is
- blank responses are coded as well, to show where the informant failed or refused to answer the question. The number used is one that will not appear in other coding, like 99, if this is well outside the range of other possible responses

Scales

The response scale for a question on a form or in an interview is usually written from least to most, with the numbers increasing accordingly. Sometimes, however, the 'least' option is the most positive response (least dissatisfied, least bullied.) It is important that all the coded scales are in the same direction (all positive or all negative) to enable data from different questions to be compared or cross-tabulated. This reordering has to be done at the coding stage rather than on the data collection form to avoid confusing informants.

Data entry

Data can be entered onto coding forms by hand or directly into a computer.

- To find and reduce coding errors:
    - have two people enter data separately then compare results (double-entry)
    - or have someone verify data entry by spot-checking for errors against the original coding form or questionnaires.

Common errors to look for are:

- numbers outside the range of codes (6 for a 1-5 response option)
- transposed numbers (1.52 instead of 1.25.)

Automatic range checks are provided by most commercial data entry programs.

6.6 Using statistics to describe data

Analytical tools

Basic data analysis can be accomplished

- with a computer software package (necessary for a large data base)
- by putting data into table form using a spreadsheet as in Excel or
- with the table function in a word processor

Other software at varying levels of complexity is available free or for purchase on the web.

Types of analysis

Analysis can take two forms. The first describes the data, shaping the results to bring out patterns and trends that may be hidden. This process is easy for non-statisticians using fairly basic mathematics and is the logical follow-up to data collection. The second more complex type of analysis applies mathematical tests to give a statistical estimate of the level of confidence one can have in the accuracy of the findings. That level of statistical analysis is only touched on here.

Descriptive statistics

- percentage (the number of responses for each option/variable as a percentage of all responses)
- mean or arithmetic average (the sum of scores divided by the number of responses to the question)
- median or mid-point response with an equal number of responses above and below it)
- mode (the most frequently occurring response)
- range (the amount data are dispersed: difference between the highest and lowest values, or range of categories with at least one response)
- standard deviation (the average of the distances between each value and the mean of all the values. See Glossary.)

How to use descriptive statistics

Most data used in descriptive statistics will either consist of

- named categories: individual items, either one thing or another (e.g., male/female or English/French/Italian)
- a choice that falls somewhere on a quantifiable spectrum of options (e.g., smaller to larger, less to more)

Categorical or nominal data

Categorical data are best described by counting how many informants' responses fall within each category (frequency distribution.) The following example shows how basic descriptive statistics can allow programs to see, and show, such patterns.

Example:

Consider the following question:

Which activity in the community recreation program do you most like participating in?

a) art activities
b) computer skills
c) music workshops
d) pottery
e) reading buddies
f) gym games

No one would ask, "what is the average gender of participants?" or "what is the average language spoken in the class?" You cannot create an average for discrete, named categories. Nominal or categorical data can best be described by counting how many informants' responses fall within each category (frequency distribution.)

Number of responses (N=23, 12 girls and 11 boys)
Responses: a = 2, b = 4, c = 5, d = 3, e = 3, f = 6

Since the responses are different from one another, but don't lend themselves to any order, counting or measurement, they are categorical (or nominal.) Calculating an average (mean) or median would be meaningless for these responses, but the program can examine the range (which responses were selected) and distribution (how frequently responses were selected) by calculating percentages or proportions (the number of times each response was given, divided by the total number.)

The results (arranged in descending order) are:
F = 6 (26.086)
C = 5 (21.739)
B = 4 (17.391)
D = 3 (13.043)
E = 3 (13.043)
A = 2 (8.695)

Although readers can determine the relative position of responses from a list like this, plotting a bar graph of percentages will make it easier to show staff differences in response levels and ask for feedback. Initial results show that f) is the modal response, (the most frequently selected option) meaning that gym is the most enjoyed activity. However, this initial analysis may raise other questions. For example, did boys and girls have different preferences? Since there are almost matching numbers of girls and boys, looking at those results could provide further information.

Ordinal data

When the options given for a question can be arranged in some order (one is bigger, better or more of something than another), it is an ordinal scale. An example would be questions with word options like: *very happy, happy, neither happy nor unhappy, unhappy, very unhappy*, which have a definite order but no equal or even definite distance from one option to the next. Because they lack a measurable, mathematical interval between them, calculating a mean or average level of happiness for the group is also not really appropriate.

Median, mode and range

Instead of mean or average, a programmer can ask about the median, the person who is in the middle of the group in terms of attitude, with half the range of responses on one side and the other half on the other side. This is useful because it shows you the trend of the responses.

Looking at the following example:

8. How do you feel about playing with other children in the recreation program?

very unhappy	unhappy	neither happy nor unhappy	happy	very happy
1	2	3	4	5

Q8, N = 15, Responses from coding sheet

playing with other children	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Totals
very happy	x	x		x			x						x		x	6
happy								x		x	x	x				4
neither																0
unhappy			x		x	x			x							4
very unhappy														x		1

To find the *median* if there are not a lot of responses

- put all responses in order then count to find the mid-point.

For this question and program, the median response is "happy.'

The mode for this data set (the most frequent response) is "very happy."

If two options had both had the most frequent number of responses, they would both be modes and the data would be called bi-modal. (If all categories have the same number of responses, there is no mode.)

Both the median and mode for these results would be encouraging for programmers.

Distribution of responses

As in this example, responses may not be evenly distributed across all the response options. Often they will be clustered at one or more typical responses, with only a few people giving quite different responses.

The pattern of distribution has more effect on the mean value than on the median. A very few responses that are quite different from the majority can skew the mean either up or down, and provide a less than accurate picture of results. Because of this, the median is usually a more useful statistic and you may want to compare all three: mean, median and mode.

The range, another useful measure, can be easily illustrated in a table or bar chart. In the example given, the majority of respondents are happy playing with others in the program. However, results show a broad range, with fully one-third of children unhappy to some extent. A narrower range of responses with no child selecting the bottom two or three response options would have been preferable. Since the data refer to a specific and small group of children, the results may provide some clues about the operation of the program, relationships with volunteers or the dynamics of that particular group. It may raise possible questions about cliques or bullying that may be more fully explained by qualitative data from observations.

Quantifiable data

The full range of descriptive statistics already described can be used for quantitative ordinal data. In this type of data, each response option is 'so many units more than another' or 'so many times more than another' on the scale being used.

Examples from community programming data collection would be:

1) quantitative questions (*how much, how many, how often?*) that provide a scale with numbers or
2) questions that ask for measurements like height, weight, or test scores.

It is common to divide data into quartiles, the responses at the 25th and 75th percentiles then plot responses on a curve. This can be done simply by first finding the median, then finding the median of each group of responses on either side. See Glossary for a more detailed formula.

These types of data can provide more precise information and can be described in more ways:

- by finding the mean response
- the median response (and quartiles)
- the mode
- range (determined by subtracting the smallest value from the largest value)
- standard deviation (a measure of variability of responses not often used in community programming evaluations)

Handling 'extreme' responses

Statisticians often treat results that lie at the extreme ends of a distribution as expendable, especially when working with large numbers of responses. In smaller-scale community evaluations, results that lie at the bottom of a distribution, reflecting

dissatisfaction or less positive results, may also provide information about program areas needing improvement. They may need to be viewed as red flags or challenges to seek further for explanations.

Cross-tabulation

Answers to the original evaluation questions are often found with simple tabulation of responses and descriptive statistics as explained above. However, there may also be a need to look at subgroups among respondents and compare results from certain questions. Cross-tabulation, which examines the relationship between responses from two questions, allows for more complex ways of looking at the data. For example, a cross-tabulation might look at the attendance/participation records of respondents in a skills development program compared to changes in skill levels before and after a program. This is an easy operation for simple statistical software, some of which can be downloaded without cost from the Internet (See References and Resources.)

6.7 Knowing about inferential statistics

As already mentioned, a second level of statistical analysis can help determine how meaningful the data are, (i.e., how well they represent reality, rather than being the result of chance) and how well they might pertain to another population.

Such statistical analysis can be done by community groups but how and when it should be done depends on a variety of factors related not just to the data, but to the whole research design, e.g., the type of data, number and selection of respondents and their characteristics compared to the larger population. Such analysis should be done by staff, volunteers or community partners with specific knowledge and experience who can help design data collection from the outset.

However, it is helpful to have some knowledge of the jargon to understand what experts are able to provide and to aid in assessing research related to similar programs or interests.

The most commonly accepted way to express the odds of some finding in the data being true (an accurate reflection of reality), rather than the result of chance, (i.e., the finding was the result of some other incidental factors) is the p-value. Individual evaluators can decide what level of p-value is appropriate for the data being investigated but generally $p<0.05$ is acceptable. This means there is only 1 chance in 20 that the result you see from the data is accidental. Any results meeting the $p<0.05$ level of probability are called 'statistically significant' and those with smaller p-values (e.g. $p<0.001$) are 'statistically highly significant.' Data without that p-value may still be accurate but need to be viewed with more caution.