

Primer for Evaluating Outcomes

Chronic Disease Self-Management Program
Kate Lorig, RN, DrPH
Diana Laurent, MPH
Stanford Patient Education Research Center
April 2007

Getting Started

There is no such thing as one questionnaire for measuring the outcomes of the CDSMP. When thinking of evaluation, every organization has to decide what is that they want to know. Before going further, let's define some terms.

Identifying information — these are questions that ask name, address, phone number, email and sometime an alternative contact person with name and phone so you can find people if they move.

Demographic information — these are questions that ask age, education, marital status, gender, ethnicity etc.

Scale — this is one or more questions (items) that measure only one thing such as pain, self-efficacy or exercise. To build a scale or even test a single question takes lots of work — in fact there is a whole field call psychometrics devoted to building and testing scales. This means that you cannot change anything on a tested item or scale.

- You **cannot pick and choose** questions. Use the whole scale or no scale.
- You **cannot change words**. Well...sometimes you can, but check with the person who developed the scale. Usually changing words invalidates the whole scale.
- You **cannot change the scaling factor (responses)**. For example you cannot change a 1-10 scale to 1-5. To give you an idea of why, just consider that to move from a 2 to a 3 on a ten-point scale the change is 10%, while moving from and 2 to a 3 on a five-point scale is 20%.

Questionnaire — this is made up of identifying information, demographic information and a series of scales. The problem is that there is an unending list of possible questions and outcomes, so there has to be some way of sorting these out. Thus, you might want to start by asking yourself these questions.

1. What do you want to know? Then for each answer ask why is this important? If you cannot give a really good answer as to importance then this is probably not something you need to evaluate. For example, you may want to know about ethnicity but do you need to know if someone is of Puerto Rican, Mexican, Cuban, or Central American origin, or is Hispanic/Latino enough? Sometimes, this is very important and sometimes not. This has to be decided at the local level.
2. Are there data that are required by the funder or by your agency? If so you will need to include these in your evaluation.
3. Are there good scales for collecting the data that you want? Most evaluators do not have the time or resources to create new evaluation scales.
4. How many questions will your participants tolerate? There are many good evaluation scales that are very long. There is always a trade off between what you want to know and what participants are willing to complete.

5. Are the scales that you are using sensitive to the type of change you are expecting? For example a weather thermometer is a fine measuring tool but is not very useful for finding a fever in a human. One or two degrees or even a half a degree makes a big difference when measuring body temperature, but not a lot of difference when deciding whether or not to wear a sweater.
6. Do you have adequate data collection resources? This means collecting data, coding it, getting it into some computer system, and most importantly following up immediately after receipt on missing data and data that looks funny. For example, the person who says he/she was hospitalized five times but reports only three days in the hospital, or someone who says they exercise for 180 minutes per day. Remember that you have to go through this whole process with every person for each follow-up. However, here it gets more complicated as you may have to contact a person several times to actually collect the follow-up data. We usually send the questionnaire by mail, then send a postcard reminder ten days later, call ten days after that and send a second questionnaire ten days after that. You may not want such a complex system, but if you are not collecting 75% or more of the data at post test, then the data you have is not very useful.

Suggested Outcome Measures

At Stanford, we usually think of outcomes in four categories; health behaviors, health status, health care utilization and self-efficacy.

The following are our favorite scales and you might want to consider them. They are certainly not the only scales and the bottom line is that the choice is yours. Unless otherwise noted the scales, their psychometric properties, and scaling are on our website at the following link: <http://patienteducation.stanford.edu/research/> Also please note that many of the scales are also available in Spanish.

Health Behaviors

Exercise measures the minutes per week of range of motion and aerobic exercise. This gives you a general idea if exercise is changing or not. There are many better scales for measuring exercise than the one we use, but most of them are very long.

Communication with Physicians. This scale may be important as it helps physicians know that patients have better communication skills with them after the workshop. Please note that you can change the word “doctor” or “physician” to “health care professional” or “provider” if you wish.

Healthy Eating. We have worked long and hard to find a good scale for measuring this. Unfortunately most scales are very long and we have never come up with a scale that meets all the criteria outlined above. If you find something you like, let us know and we will put it on our web site and share it.

Health Status

Self-rated Health. This is a single item and is very strong at predicting future health. It is part of the National Health Survey and thus can be used to compare with other data.

Fatigue Visual Numeric Scale. Again, a single item. Fatigue is the most common symptom across chronic disease and should be one of your top choices. The Visual Numeric Scale does take up paper space but is very easy for people to use—it is much better than the more traditional visual analogue scale. We have used it with low literacy groups, Hispanics, and older people.

Pain Visual Numeric Scale. Not all people with chronic disease have pain but since arthritis is the most common disease in the elderly, most people have some pain. All the comments about the fatigue scale apply here.

Shortness of Breath Visual Numeric Scale. Shortness of breath is less common than pain or fatigue, but having said this we are surprised that we usually see changes in this scale even with people whose primary diagnosis is diabetes. The same comments apply as for the fatigue scale.

Health Assessment Questionnaire (HAQ) Disability Scale (8 items). This scale is also used in the National Health Survey and is good at measuring physical disability. The questions are designed to measure the function of the major joints in the body, not activity of daily living. If there is a trade off because of length between HAQ and a depression scale, we would go with depression.

Health Distress. This is a small four item scale to measure how worried or distressed one is because of health problems. It is not a true depression scale but seems to work well across language and culture groups which cannot be said of many depression scales.

Social/ Role Activity Limitation Scale. Again, a short scale that measures role function. This is one of the prime movers of quality of life for most people and we would be inclined to always include it.

Depression. This is a hard one. Most depression scales such as the Beck and the Zung only score for clinical depression. They are not sensitive in the area of subclinical depression which is the problem for most people with chronic illness. The CES-D has long been used as it is pretty good for picking up subclinical depression. The problem is that it does not change with cultures and languages well. Right now our favorite depression scale is the Patient Health Questionnaire 9 (PHQ-9). This is not on our website but you can find it at this link:
<http://www.pfizer.com/pfizer/download/do/phq-9.pdf>

Health Care Utilization

These questions are very important as it is this data that causes policy change. It is also the data that can be used to compare costs to effects. We suggest that you use four measures.

- **Outpatient Visits to MDs**
- **Visits to Emergency Departments**
- **Number of Hospitalizations**
- **Nights in Hospital**

While it would be really nice to know about visits to nurse practitioner, chiropractors, etc., the public just cannot differentiate all these different people, and so after long hard experience we just ask about MD visits.

We ask about nights in hospital rather than days because people often confuse a day surgery with a hospitalization. If they were there overnight then they really were hospitalized.

This is certainly not a complete list of utilization but does give an idea of what is happening. These questions should always be asked based on the time between questionnaires. Thus, if you have a baseline and a six month questionnaire, you ask about visits in the past six months. If you have a baseline and 4 month questionnaire you ask about visits in the past 4 months. You need to be sure that the timing on these questionnaires is the same for all questionnaires and that the visits do not overlap. That is you cannot ask about visits in the past four months before the class and then at the end of the class again ask about visits in the past four months because only 2 months will have past since you asked the question the last time.

Self-efficacy

Chronic Disease Self-efficacy Scale (6 items). Please note that self-efficacy is a weak outcome. No one wants to go to their state legislature and say this is a great program because it increases self-efficacy. If this is the only outcome you get, then there is not much there. Self-efficacy is important as it predicts changes in health behaviors and health status. The people who tend to do best in the program as those who either have pretty good efficacy to start with or who improve their self-efficacy with the workshop.

Formatting Questionnaires

There is an art to formatting questionnaires, which comes from coding thousands of questionnaires and seeing response patterns emerge from our mistakes. Here are the suggestions that we have learned over the years.

Paper

- The easiest colors on the eye are white, ivory, beige, and light yellow. Avoid bright colors or dark colors. At Stanford, we color code our studies by the color of the questionnaire, so we've expanded to light blue, light green, and pink, but we have several studies going at any one time
- Having a heavier paper is nice, but it costs more – both to print and to mail.
- The most professional-looking questionnaires are in booklet form, printed on 11" x 17" paper, folded in half and stapled at the crease. Pages are less likely to come loose than with a questionnaire of 8½" x 11" paper stapled in the corner. If you use a booklet format, however, you questionnaire must have pages in multiples of 4 (4, 8, 12, 16, etc.). You can fill extra blank pages with a cover letter or thank you note, if you are short.

Type Size and Style

- Use 12-point type for the questions and instructions. If you are working with a population where eyesight may be an issue, use 14-point type. Avoid type smaller than 12-point.
- Do not use fancy type; stick with Times Roman or Arial. Times Roman takes less space than Arial for the same size type.
- If necessary, 10-point bold type can be used for scale anchor words (e.g., “not at all confident”, “totally confident” on self-efficacy scales) or response column headings.

- Bold or italics can be used for emphasis in a question, for example: “How many **days** in the **last 6 months** did you...”
- Generally, avoid colored type. Keep it simple.

Spacing

- Do not cram your questions together to save paper. There should be space between the question and the responses, and there should be additional space between questions. The closer you questions are to each other, the more likely that questions will be missed and left blank.
- Allow adequate page margins, especially at the top and bottom of the page. If you have a question close to the top of the page, the person’s eye may skip over it and it will be left blank.
- If you have a multi-item scale where the question is on the left and the response categories to be circled or checked are spaced out on the right of each question, the eye has trouble scanning across to the correct place. Sometimes, they will circle the response for a question below or above the correct one. To help avoid this, there are two things you can do: 1) set the first response tab stop to have leading dots between the question and the first response category, 2) place a gray line between every second question (can be the entire blank line shaded in Word).

Organizing the Scales

- Generally, the timing for the questions should be consistent for all questionnaires and all questions. For example, changing from 2 months to 4 to 6 will confuse people. Almost all of Stanford questionnaires are asked for the time period between questionnaires. There are exceptions, however. For example, there are some questions (disability, pain), where a person may not be able to judge over a long period, and you may ask for the past 2 weeks (pain) or now (disability). If your question or scale has a different time requirement than the rest of the questionnaire, bold the time period in the question.
- The direction of the responses should be consistent throughout the questionnaire, for example “none of the time” on the left and “all of the time” on the right, and “not at all confident” on the left and “very confident” on the right. If you switch response direction, you may lose your data because people aren’t paying attention.
- Break up the scales into logical categories with prominent headings, such as: Background Information, Symptoms, Medical Care, Feelings, Physical Abilities, etc.
- Avoid breaking a scale over more than one page, if possible. This is especially true if one question of the scale is orphaned on a different page.
- If the scale is too long to fit on one page and you have to break it, the second page should have at least 2 questions from the scale on it, and not be too close to the top of the page. Repeat the question and the descriptions of the response columns (“never”, “seldom”, “sometimes”, etc.) on the second page.
- Number your questions. If the questions are more than one part, then the second level should be numbered with a, b, c, d, etc.

- Scales or questions that may be uncomfortable for people to answer should be placed with non-threatening, easy scales around them. Depression should not be next to disability, for example. Mix them up.
- Difficult questions should never be first. People may abandon the questionnaire. Sneak up on them later in the questionnaire.
- Questions that ask information that may require that the person look something up (such as medical visits or medications) should be at the end of the questionnaire. If people have to leave the questionnaire early, they may not return to it. If they invested time already, they are more likely to finish.

Coding Bars

If you use Word to create your questionnaire, you may find yourself swearing at your computer when it comes to the coding bars. The coding bars are the columns on the outside of each page where you have the variable names and a line to record the codes for each question in preparation for data entry. We've created coding bars two ways: either by creating columns in Word, or text boxes. The text boxes seem to work the best. Columns are a nightmare because they have to switch from the right side on odd numbered pages to the left side on even numbered pages. If you change something in the slightest, the columns are messed up.

Text boxes can be created and the coding column created inside it. They can be resized more easily. If you size and place the text box so that three of the sides are outside the margins (and won't print), you will have a ready-made line defining the coding column on the one side that's left in the printable area. They tend to jump around if you make changes in your questionnaire, which is irritating, but it's easier than columns.

Coding bars can be formatted in these ways:

- We use 10-point Arial Narrow for the coding bars. This takes up less room and allows more space for the coders to write.
- You can shade the coding bar gray if you wish
- The top of each column should say "LEAVE THIS AREA BLANK" in bold. It can be 2 lines.
- The coding line for the question should be as close to directly in line with its corresponding question as possible.
- If you have quality control system in place for your data, the end of the questionnaire should have places in the coding bar for staff to initial that they have complete a task. At Stanford, we have lines labeled: coded (person who coded questionnaire), called (person who made any follow up calls for missing or unclear data), checked (person who checked the coding for accuracy), entered (person who entered the questionnaire into the database), verified (person who checked the entry for accuracy).

Footers

- Footers on each page of the questionnaire should contain the page number on the outside of the page (right side on odd pages, left on even) – even better, it could say page "1 of 6", "2 of 6", etc.
- You can add the name of the study on the opposite side of the footer, if you like.

- Footer type should be slightly smaller than the rest of the questionnaire, so the eye doesn't perceive it as another question. We use 10-point type for our footers.
- In Word, you can highlight the footer and go to Format: Paragraph and set the Spacing: Before to "6". This will give the footer an extra half-line space between it and any text above.

Stanford Patient Education Research Center
1000 Welch Road, Suite 204 • Palo Alto CA 94304
(650) 723-7935 • (650) 725-9422 fax
self-management@stanford.edu
<http://patienteducation.stanford.edu>