
PLCO SUPPLEMENTAL QUESTIONNAIRE USER GUIDE

INTRODUCTION

The Supplemental Questionnaire (SQX) was introduced in 2006, 13 years into the trial. The contents of this questionnaire largely overlap with information collected on the baseline questionnaire (BQ). Some of the overlapping information is data that would not be expected to change between the BQ and the SQX (e.g. race-ethnicity). Other variables may reasonably change between the two questionnaires (e.g. marital status, smoking status, diseases). The SQX additionally contains some data not collected on the BQ (e.g., physical activity).

Ultimately, about 104,000 questionnaires were collected. Given the late administration of the SQX, the median time into the trial this form was completed was 9 years and the median time following a completion of this form was 3 years.

Customized datasets of only selected variables can be constructed upon request on a case by case basis.

COMPLETION

All PLCO participants are included in this dataset. The variables SQX_STATUS and SQX_SUBSTATUS describe who was expected and the resulting compliance. Over 103,000 valid questionnaires were completed and are available. The valid questionnaires are identified by the variable SQX_VALID and require the following criteria:

- A date of completion must be available.
- The date of completion must be prior to the date of death.
- At least one leading non-verbatim question that is neither 'mark all that apply' nor gender related must be answered.
- The participant answered the correct gender questions only and none of the other gender questions.

NAMING CONVENTION

There are three prefixes used to name the supplemental questionnaire (SQX) variable names. There are two sets of variables collecting information only from the SQX and one set of variables that compare the SQX answers to the baseline questionnaire (BQ) answers. The SQX and BQ information was self-reported; the PLCO trial makes no effort to verify the validity of participants' answers.

- **SQX_** These questions were asked on the SQX only and not asked on the BQ.
- **SQXO_** These questions were asked on both the SQX and BQ. Only the SQX answer is used for this version.
- **SQXBQ_** These questions were asked on both the SQX and BQ and both answers are compared. Variables were set to ambiguous if the SQX contradicts what was previously answered in the BQ.

MISSING CODES

Missing codes have been standardized to represent the same thing for all variables across all datasets. They are as follows:

| CODE | FORMAT | FURTHER EXPLANATION |
|------|----------------------|--|
| .A | Ambiguous | Contradictory information resulted in an unclear conclusion. |
| .F | No Form | The SQX form was not filled out. |
| .G | Wrong Gender | The variable only pertains to one gender. |
| .M | Missing/Not Answered | The question was not answered or data are otherwise missing. |
| .N | Not Applicable | The variable did not apply to this participant. |
| .R | Out of Range | The value was considered either too extreme or impossible. |

AMBIGUOUS ANSWERS

There are two reasons for ambiguous answers:

- The participant marked more than one answer.
- If a positive response on the BQ is contradicted by a negative response on the SQX then this version of the variable is set to ambiguous. For example, this includes people who report a heart attack on the BQ and not on the SQX.

TIME AT ENTRY

Entry for the SQX starts when the SQX was completed, after randomization and the baseline questionnaire (BQ). `SQX_DAYS` identifies how many days after randomization the SQX was completed and can be compared with other time events in PLCO.

HEALTH AT QUESTIONNAIRE

For some analysis, participants diagnosed with cancers are excluded. The variable `PH_ANY_SQX` can be used to identify participants who were diagnosed with cancer prior to completion of the SQX. Self reported co-morbidity and disease variables are available in the Health History section of the data dictionary.