

International Lung Cancer Consortium: Instruction of Data Submission for the
Family History Project

In support of the pooled analyses, we ask that you send the files described to Dr. Eric Duell
(duelle@iarc.fr)

I. **The study questionnaire.** This is to ensure that the variables are comparable, and to facilitate data checking. If not stated in the questionnaire, how was “relative” defined (included grandparents, cousins, aunts, uncles)? It is assumed that “first degree” relative includes parents, full siblings, and offspring. Please let us know if the data set includes both full and half sibs.

II. **A data file** containing the following variables for each case or control:

Proband Variables:

1. ID number
2. Case/control status
3. Center
4. ICD code (please let us know what edition—9, 10, etc.)
5. Histology code
6. Control diseases (for hospital based controls)
7. Sex
8. Race/ethnicity
9. Age at diagnosis/interview (Date of birth, date of interview, date of diagnosis. If dates are not available, please provide the year of birth, interview and diagnosis.)
10. Cigarette Smoking status (ever/never/ex/current)
11. Pack years of smoking
12. Pipe smoking (yes/no)
13. Years of pipe smoking
14. Cigar smoking (yes/no)
15. Years of cigar smoking
16. Education
17. Social economic status

Family History Variables:

1. Lung cancer in mother (yes/no)
2. Lung cancer in father (yes/no)
3. Lung cancer in full siblings (yes/no)
4. Number of full siblings with lung cancer
5. Number of full siblings total (with and without lung cancer)
6. Lung cancer in offspring (yes/no)
7. Number of offspring with lung cancer
8. Number of offspring total (with and without lung cancer)
9. Other cancer(s) in mother (yes/no)
10. Cancer type(s) in mother (ICD code)
11. Other cancer(s) in father (yes/no)
12. Cancer type(s) in father (ICD code)
13. Other cancer(s) in siblings (yes/no)
14. Cancer type(s) in siblings (ICD code)
15. Other cancer in offspring (yes/no)
16. Cancer type(s) in offspring (ICD code)
17. Number of relatives with lung cancer
18. Number of relatives with any cancer
19. Total number of first-degree relatives (with and without cancer)
20. Total number of relatives (both with and without cancer)

Please include, *in the same data file as the case/control data*, the following information for each relative, if available:

Relative Data:

1. Family ID (matching the case/control ID)
2. Individual ID
3. Family member of case or control
4. Relationship to case or control (mother, father, sister, brother, etc.)
5. Sex
6. Vital status
7. Age at interview or death
8. Lung cancer diagnosis (yes/no)
9. Age at lung cancer diagnosis
10. Histologic type of lung cancer
11. Other cancer(s) diagnosis (yes/no)
12. Type of other cancer(s)-(ICD codes)
13. Age(s) at other cancer diagnosis
14. Smoking status (ever/never/current/ex/)
15. Packs per day
16. Years of smoking
17. Pack-years

All personal identifiers should be removed, but please keep the ID number in order to facilitate queries on the data.

File format: A SAS file is preferable, but Excel or ASCII file formats are also acceptable if they are more convenient for the sender.

III. A data dictionary listing:

- a. the names of each variable being sent
e.g., tob_py, fam_his, etc.
- b. a brief description of each variable
e.g. number of cigarette smoked per day;
e.g. relatives who had cancer ; the cancer site of the affected relatives;
- c. definitions of data codes used for the variables
e.g. 0=never, 1=ever
e.g. 1= father; 2=mother; 3=brother; 4=sister; 5=son; 6=daughter

Also, if there are any specific issues associated with the use of particular variables, it would be helpful to include comments describing these issues. (This is optional.)