The French National Survey on Indoor Air Quality: sample survey design

F. Golliot[a,*], I. Annesi-Maesano[b], M.C. Delmas[c], F. Dor[c], Y. Le Moullec[d], L. Mosqueron[e], V. Nedellec[e], J. Ribéron[a], G. Salines[c], S. Kirchner[a]

[a]*Centre Scientifique et Technique du Bâtiment, Marne la Vallée, France;* [b]*Institut National de la Santé et de la Recherche Médicale, Paris, France;* [c]*Institut National de Veille Sanitaire, Saint Maurice, France;* [d]*Laboratoire d'Hygiène de la Ville de Paris, France;* [e]*Vincent Nédellec Consultants, Paris, France*

ABSTRACT

A nationwide survey on indoor air quality will be set up in France in 2003–2004. The aim of the survey is to assess household exposure to indoor air pollutants. The target population is the national housing stock of approximately 24 million permanently occupied housing units. A nationally representative sample of dwellings will be selected based on a three-stage sampling design. At the first stage, the Primary Sampling Units (PSUs) are categorized based on French territorial divisions. The second stage of the design is the sampling of segments within the PSUs. The third stage consists in registering eligible dwellings within each segment.

In such a design, each housing unit has an equal probability of selection.

INDEX TERMS

Sample survey design; Indoor air; Air quality

INTRODUCTION

The first nationwide survey on indoor air quality will be set up in France from June 2003 to April 2004. The French building research centre (CSTB), which is in charge of setting up and conducting the survey, has organized the project as a network of scientists working in the fields of public health, chemistry, microbiology, physic, environmental, building and social sciences.

The main objective of the survey is to assess indoor air quality in French dwellings and to identify its determinants. This survey will provide useful data for risk assessment and risk management related to exposure to indoor air pollution and will allow to better target preventive policies (technical management policies, development of new products and techniques in the building field, public health policies). Household members will be interviewed using standardized questionnaires. In addition, descriptive data on the homes and environmental samples will be collected. Data collected from the survey sample will be used to assess the overall population exposure. The survey has to provide reliable information about distributions of pollutants concentrations, buildings characteristics and time activity patterns of households, in order to be used in risk management.

Bias in the sampling procedure could overestimate a specific type of building or a specific population of households and distort the risk assessment, which could lead to flaws in the strategy of risk management. To avoid this, the household selection method was strictly defined, taking into account the scientific objectives together with the financial and organizational constraints of the study.

The aim of this article is to describe the sampling strategy that was chosen to insure an objective selection of housing in the French nationwide survey on indoor air quality. This sampling procedure is still under progress.

---

* Corresponding author. E-mail: franck.golliot@cstb.fr

METHODS

Constraints and Limitations

This project is limited by various constraints.

- The cost of the survey for each housing unit limits the number of units to be investigated.
- The dwellings must be occupied at the time of the survey.
- The sample variability must be as close as possible to the variability of the national housing stock of permanently occupied housing units in 2003, in terms of building characteristics.
- Sampling units must be concentrated in geographical areas in order to make the organization easier (quantity of material needed and number of technicians).
- The sampling procedure must be unbiased in order to make correct statistical inferences. The large number of pollutants to measure, the poor knowledge of concentration levels and the large variability of dwellings and households demands that the sample selection be according to a random method to ensure that units are drawn from everywhere in France.

Target Population

Since the objective of the study is assessing indoor air quality in dwellings, it was decided that the target population should be the French housing stock of permanently occupied housing units. This stock is about 24 million at the time of the study (from June 2003 to April 2004). Vacant dwellings, secondary homes, hotels and institutional housings such as old people's homes, convents and jailhouses were excluded. Due to organizational constraints, the target population was also restricted to mainland France, excluding overseas territories.

Data to be Collected and Parameters to be Estimated

The principle of statistical inference is to estimate characteristics of a population (e.g. means, standard deviations, proportions) on the basis of information collected from a sample.

In the present study, various data, such as concentrations of chemical compounds (volatile organic compounds, carbon monoxide, particles), concentrations of biological compounds (cat and dog allergens, mites, bacteria, moulds), physical parameters (temperature, relative humidity, air renewal), counts (rooms, people, hours spent indoors), and nominal data will be recorded. According to the data collected, various parameters will be calculated. The most often used are: the central tendency (arithmetic and/or geometric mean), the variance and the percentiles of distribution for measurements and the frequencies for nominal variables.

Sample Size

The determination of the sample size is a question that arises early in the planning of any survey. Samples larger than necessary constitute a waste of resources whereas samples too small provide inaccurate data that are useless for decision making. The main objective is to obtain both a desirable accuracy (or bearable uncertainty) and a desirable confidence level with minimum cost.

For a large population, the accuracy is determined only by sample size for a given degree of confidence (a smaller risk of error requires a larger sample for the same level of uncertainty). To estimate the number of housing units needed, results from the pilot survey conducted in 2001 in 90 dwellings were used (Kirchner *et al*., 2002). Although the sample was not drawn randomly, the results of the pilot study gave some indication on the variability of various compounds. The sample size was determined according to four criteria: (1) the uncertainty accepted for the estimates; (2) the compounds for which it is essential to have a precise

estimate (the answer to this question was brought by a specific work on the scoring of substances based on risk assessment (Mosqueron, 2003); (3) the degree of confidence; and (4) the available resources.

It was calculated that at least 500 housing units should be included in the study in order to provide estimates with an uncertainty of less than 10% for the majority of the studied compounds and a degree of confidence of 5% (Figure 1).

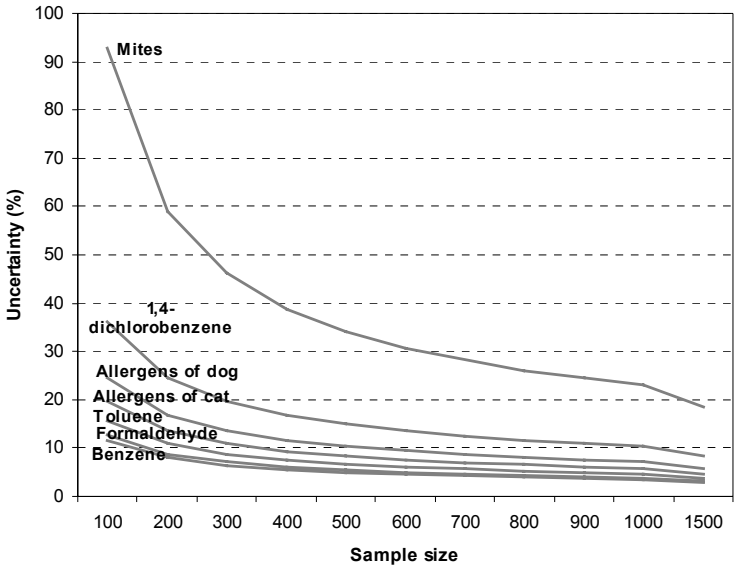

Figure 1 Estimation of uncertainty according to sample size and parameters surveyed.

However, in order to take into account failure of measurements in some housing units and a potential underestimation of the variance in the pilot survey, it was decided to draw 710 units. This sample size should provide an accuracy of estimates acceptable for risk assessment and should allow the testing of several assumptions on potential sources of indoor air pollution.

Sampling Method

For practical reasons, it was decided that sampling units should be concentrated within a few geographical areas. Housing units were therefore drawn according to a multistage sampling design. In order to ensure an equal probability for each housing unit to be selected, a Probability Proportional to Size (PPS) sampling design was chosen.

Principle of PPS sampling with three stages

A multistage sampling method based on PPS sampling at each stage (except the last one) makes it possible to select a given number of units at a given stage, concentrating units in some areas and ensuring the same probability for each final sampling unit to be selected (Cochran, 1977).

At the first stage of the design, the probability of selection of a $PSU_{(i)}$ (P_i) is proportional to its size and could be defined as follows:

$$P_i = \text{Number of PSU to draw} \times \frac{\text{Number of housing units in the PSU}_{(i)}}{\text{Total number of housing units in the population}} \quad (1)$$

At the second stage of the design, a constant number of segments are sampled within the PSU drawn at the first stage. Each segment is selected with a probability ($P_{j/i}$) proportional to its size

$$P_{j/i} = \text{Number of segments to draw in each PSU} \times \frac{\text{Number of housing units in the segment}_{(j)}}{\text{Number of housing units in the PSU}_{(i)}}$$

(2)

At the last stage, housing units are selected according to a simple random sampling in each segment. This leads to draw a constant number of housing units whatever the size of the segment. At this stage, a unit's probability of selection ($P_{k/i,j}$) is defined as follows:

$$P_{k/i,j} = \frac{\text{Number of housing units to draw in each segment}}{\text{Number of housing units in the segment}_{(j)}} \quad (3)$$

The overall probability of selection of a housing unit (P_T) is obtained by multiplying the probability of selection at each stage: $P_T = (1) \times (2) \times (3)$. This probability is equal to the sampling rate, which is the size of the sample divided by the population size.

Implementation of the sampling method

The first stage of the design consisted in selecting the Primary Sampling Units (PSU) by sampling Smallest Territorial Divisions (STD). Mainland France is divided into 36 201 urban or rural STDs (i.e. towns or villages). Adjacent STDs with less than 50 dwellings were grouped into a single PSU. The 36 201 STD were grouped into 32 105 PSUs. All PSUs with more than 100 000 dwellings corresponding to the 11 largest French cities were included (i.e. probability of inclusion of 1). Among PSUs with less than 100 000 dwellings, 63 were sampled with a probability proportional to the number of dwellings. PSU sampling was stratified according to eight geographical areas and four urbanization levels (rural, small, medium, and large). The urbanization levels are defined by the French National Institute of Statistic (INSEE).

The second stage of the design corresponded to the sampling of segments within each PSU. In the local land registers, STDs are divided in segments corresponding to groups of streets or of districts. Ten segments were selected within each PSU with less than 100 000 dwellings. For PSUs included with certainty (i.e. PSUs with more than 100 000 dwellings), the number of housing units to draw in each PSU was calculated in order to ensure that everyone of them has the same probability to be selected as the housing units from PSU of less than 100 000 dwellings. A total of 710 segments were selected with a probability proportional to their sizes: 80 segments in PSU included with certainty and 630 in PSU drawn with a probability smaller than one.

The third stage will consist in selecting housing units within the selected segments. In some PSUs, the local land registers are computerized. In these cases, a sample random will be drawn easily. In PSUs without electronic land register but of moderate size, each housing unit will be registered by visiting the segment. In large PSUs without electronic register, the phone directory will be used to list eligible housing units. Only one housing unit per segment will be surveyed in order to avoid making measurements in neighbouring housing units in which the variability is likely to be smaller than between distant units.

The number of permanently occupied housings within PSUs and within segments was obtained from the 2001 National Census of Housings (FILOCOM). This file is built by the taxation authorities from the inhabited house duty, the land tax and the income tax. Due to the sensitive contents of the file, no address can be used.

Screening and Household Recruitment

Based on previous survey describing homes and taking environmental samples, a response rate from 20 to 50% is expected (Leech *et al.*, 1996; Seifert *et al.*, 2000; Vojta *et al.*, 2002). To finally select 710 households, five households per segment will be contacted, using an equal probability sampling. This part of the survey (assistance to final selection and contact with eligible households) was subcontracted to the French Research Centre for the

Observation of Living Conditions (CREDOC). Before the interviewer's visit, each of the sampled households will receive a letter explaining the purpose of the study and informing the householders that an interviewer would be visiting soon. The interviewer will attempt to contact an adult resident at each housing unit, up to three times on different days of the week and hours of the day. After making contact with an adult householder, the interviewer will administer a short questionnaire describing building and household characteristics and, for people who do not want to participate in the survey, the reasons of refusal.

RESULTS AND DISCUSSION

The first French National Survey on Indoor Air Quality aims at gathering data on indoor air pollutants and building characteristics from a nationally representative sample of the French housing stock. Because bias in the sampling procedure could overestimate a specific type of building or a specific population of households, it was decided to select housing units using a random method ensuring to draw units objectively everywhere in France.

A design based on a three-stage PPS sampling has concentrated selected housing units on 50 of the 94 France subdivisions, and enables to survey at least 10 housings in each PSU (Figure 2). This reduces the distances between housing units to survey and thus the survey cost. The sampling of segments within PSUs, allows us to register housing units within eligible segments before selecting housing units. Due to the existence of a cluster design effect, multistage sampling is less accurate than simple random sampling (Cochran, 1977). In such design housing units from a same PSU tend to be more alike in their characteristics (seasons, building structure, environment, etc.) than distant units. To reduce this effect, we selected only one housing unit per segment in the second stage of the design. Although our method is a complex sampling design, it ensures the same final probability of selection for each housing units, making easier statistical inferences.

Because the aim of the survey is to assess indoor air quality, we defined the target population to the dwellings rather than individuals. This would lead to a sample representative of the household (a household is attached to a dwelling), but not of the individuals. Within a household, individuals tend to be more alike in their characteristics and behaviours than individuals randomly selected in the population. To assess individual exposure, we will collect data on each occupant and we will ask each one to fill out a time activity diary. The non-independence of occupants within the household should not be ignored for data analysis.

The household recruitment is a critical point in such a survey. Because a poor response rate could lead to recruitment of a specific population, it was decided to work with sociologists with expertise in conducting large national surveys on households. To prevent a poor response rate and to select finally 710 households, five households will be drawn per segment. Before making inferences, we will have probably to consider a non-response adjustment factor. Such factor inflates the base weights of the surveyed housings units so that they also represent the eligible housing units that did not participate in the survey (Särndal, 1992; Vojta, 2002). The participation rate in the study, however, remains a key issue.

Thanks to this study, data on indoor air quality in France will be available, and could be used as reference for comparisons with measurements performed in the homes of specific population such as asthmatic people. However, data collected will also be used for risk assessment (and, eventually, risk management), which necessitate to estimate exposure of people.

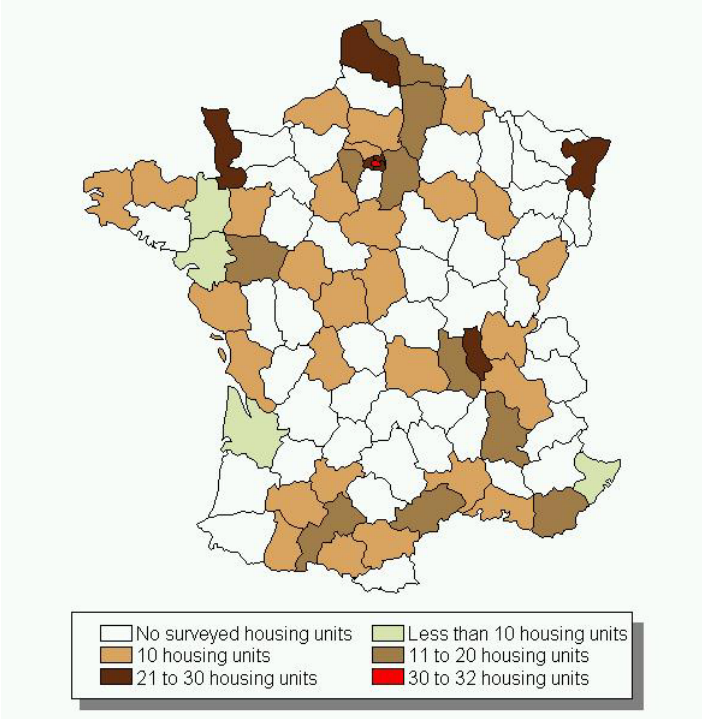

Figure 2 Map of the 710 housing units according to their localization

ACKNOWLEDGMENTS

The French permanent survey on indoor air quality is funded by the Ministries of Housing, Health, Environment, the Energy Control and Environment Agency (ADEME) and the French Research Building Centre (CSTB). We acknowledge the additional experts of the working group on sampling strategy in the French National Survey on Indoor Air Quality: J.F. Arenes (CSTB), B. Collignan (CSTB), F. De Blay (CHU Strasbourg), M. Derbez (CSTB), H. Desqueyroux (ADEME), P. Desqueyroux (Géocible), B. Festy (APPA/OQAI), C. François (CSTB), M.C. Lemaire (ADEME), J. Manel (CHU Nancy).

REFERENCES

Cochran, W.G. (1977). *Sampling Techniques*, 3rd edn. New York: John Wiley & Sons, Inc.

Kirchner, S., Pasquier, N., Cretier, D. *et al*. (2002). The French permanent survey on indoor air quality: survey design in dwellings and schools. *Proceedings of the 9th International Conference on Indoor air Quality and Climate—Indoor Air 2002*, Monterey, Vol. 4, pp. 449–454.

Leech, J.A., Wilby, K., McMullen, E. *et al.* (1996). The Canadian Human Activity Pattern Survey: report of methods and population surveyed. *Chronic Diseases in Canada* **17**, 118–123.

Mosqueron, L., Kirchner, S. and Nedellec, V. Ranking indoor pollutants according to their potential health effect, for action priorities and costs optimization in the French Permanent Survey on Indoor Air Quality. *Healthy Building 2003* (submitted for publication).

Seifert, B., Becker, K., Hoffmann, K. *et al*. (2000). The German Environmental Survey 1990/1992 (GerES II): a representative population study. *Journal of Exposure Anaysis andl Environmental Epidemiology* **10**, 103–114.

Särndal, C.E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Spinger-Verlag.

Vojta, P.J., Friedman, W., Marker, D.A. *et al*. (2002). First national survey of lead and allergens in housing: survey design and methods for the allergen and endotoxin components. *Environmental Health Perspectives* **110**, 527–532.