Proceedings of Statistics Canada Symposium 2014
Beyond traditional survey taking: adapting to a changing world

Organization of data collection and matching for the epidemiological surveillance of workers who may be exposed to nanomaterials in France (EpiNano)

Delphine Jezewski-Serra[1], Laurène Delabre[1], Stéphane Ducamp[1, 2], Yuriko Iwatsubo[1], Irina Guseva-Canu[1]

Abstract

The purpose of the EpiNano program is to monitor workers who may be exposed to intentionally produced nanomaterials in France. This program is based both on industrial hygiene data collected in businesses for the purpose of gauging exposure to nanomaterials at workstations and on data from self-administered questionnaires completed by participants. These data will subsequently be matched with health data from national medical-administrative databases (passive monitoring of health events). Follow-up questionnaires will be sent regularly to participants. This paper describes the arrangements for optimizing data collection and matching.

Keywords: Nanomaterial, data collection, matching

1. Introduction

The European Commission defines a nanomaterial as any material containing loose particles in the form of an aggregate or agglomerate, in which at least 50% of the particles, in numeric distribution by size, have one or more external dimensions between 1 nanometre and 100 nanometres (European Union 2011), whether intentionally produced or not.

Because of their properties, which give added value to the products that contain them, nanomaterials are now found in many products available on the market, even though their potential toxicity is unknown. The potential health risks of nanomaterials have not been clearly identified.

In 2007, the French directorate general of health and the directorate general of labour tasked the Institut de veille sanitaire (InVS) [the institute for public health surveillance] with developing a surveillance program on the health effects of occupational exposure to intentionally produced nanomaterials. A feasibility study (Boutou-Kempf 2011) set out the surveillance methods to be implemented. This surveillance program, known as EpiNano, is based on the establishment in France of a registry of potentially exposed workers, on general prospective monitoring (passive and active) of these workers, and possibly on medical and/or biological surveillance. Initially, the registry is limited to workers who may be exposed to nanometric titanium dioxide (TiO_2) and carbon nanotubes (CNTs) (Canu et al. 2013). The purpose of the program is to perform general medium- and long-term monitoring of the possible health effects of occupational exposure to manufactured nanomaterials, i.e., assess the risks associated with human exposure to intentionally produced nanomaterials. The Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé (CCTIRS) [the advisory committee on information processing for health research] and the Commission nationale de l'informatique et des libertés (CNIL) [the national commission on information technology and liberties] approved the scientific protocol for EpiNano in late 2013, thus authorizing the implementation of the program. Visits to businesses and recruitment of participants began in early 2014.

This paper describes the methods the InVS used to collect the various types of data and the matching techniques applied in the EpiNano program.

[1]Institut de veille sanitaire, 12 rue du Val d'Osne, 94415 Saint-Maurice Cedex, France, epinano@invs.sante.fr.
[2]Université de Bordeaux-Ségalen, Laboratoire santé travail environnement (LSTE), 33076 Bordeaux, France.

2. Data collection methods

2.1 Collection of exposure data

The program, as designed, begins with the identification of businesses that produce or use nanomaterials, with the goal of assessing the nature of the exposure in those businesses (Canu et al. 2013). Once the businesses have been identified, the InVS arranges a visit to their facilities to identify workstations that could result in exposure, to characterize the exposure, and to collect the contact information of the people who work at those workstations. This visit also provides an opportunity to inform workers of the goals and approach of the program and the conditions under which they can participate. An industrial hygienist and an epidemiologist from the InVS describe and characterize the nature of the exposure on the basis of the data they collect during the visit.

These exposure data, collected with an on-the-job technical observation booklet (TOB) (Guseva-Canu et al. 2015), will be used to classify workstations as potentially exposure-causing or not. They will also be used to provide a semi-quantitative assessment of the exposure.

To simplify everyone's job, an online questionnaire with the same structure as the technical observation booklet was developed with a special software program called Voozanoo®. This is a tool for designing and creating online information collection and management systems. The hygienists can now capture data live at the visit site or when they return from their mission, depending on the technical resources at their disposal (e.g., Wi-Fi, 3G connection, laptop, or tablet). The workstation description forms are structured so that they can be attached to the visit report. Thus, businesses receive not only a conclusive report on whether the workstations are potentially exposure-causing or not, but also a summary of the information compiled by the InVS teams, so that they can take measures to protect their employees if they wish. In addition, the use of an online questionnaire makes it possible to develop a thesaurus for a number of variables before data capture, thereby avoiding the after-the-fact classification of data from write-in areas. For example, a workstation where people go to weigh materials containing carbon nanotubes might be identified by the person responsible for data capture as a 'weigh station' or as a 'precision scale'. The introduction of a thesaurus will ensure that this workstation is uniquely identified as a 'weigh station'. Aside from the problems associated with the use of uppercase letters, lowercase letters or accents, using a thesaurus to standardize data capture ultimately simplifies data classification and eliminates the need for data-grouping steps, which can result in errors when large quantities of data are captured. Consistency edits are applied to key variables from the data capture stage onward, and the data are formatted directly so that they can be used by statistical software packages (SAS and Stata in this case).

2.2 Registry of employees who may be exposed to nanomaterials

Following the visit and the transmission of the report, the business sends the contact information of the people who work at these workstations by secure means to the InVS, which takes care of sending out the recruitment questionnaires.

At this stage of the program, two types of data are collected. Industrial hygiene data are collected during the visit to the business and are used to characterize the exposure potential of each workstation. Personal data are collected using the recruitment questionnaire completed by program participants. Within a business, a given individual may be required to work at various workstations. For example, a person may go to the weigh station to weigh the powder containing synthesized raw carbon nanotubes and to the extraction station to dissolve the composite material with an acid treatment to extract the purified nanotubes. To link these two pieces of information, it was necessary to assign to each workstation a unique number used only by the InVS (composed of the business number, the shop number and the workstation number), so that each individual can identify the workstations that he or she uses. At the end, an individual exposure score reflecting the various operations performed at each exposure-causing workstation will be calculated.

The recruitment questionnaires also collect the employment histories of program participants. When the industrial hygienists receive the questionnaires, they code each job by occupation (under the 2008 French classification of occupations and socio-occupational categories) and industry (under the 2000 and 2008 nomenclature of French industries). The occupation data from EpiNano can then be cross-classified with national occupation–exposure matrices (Févotte et al. 2006) to identify the substances that people may have been exposed to during their careers. With this information, the effect of certain co-exposures can be studied in relation to the exposure to nanomaterials, and statistical models may be adjusted to control for confusion bias or to account for an interaction effect between the two exposures. For example, past exposure to respirable crystalline silica dust

must be taken into account when the relationship between exposure to nanomaterials and lung cancer is studied, because respirable crystalline silica dust is known to be associated with an increased risk of developing lung cancer (Pelucchi et al. 2006).

2.3 Medium- and long-term monitoring of participants

Medium- and long-term monitoring will be broken down into several steps, with active monitoring through regular follow-up questionnaires on the one hand and passive monitoring through available medical–administrative databases on the other.

With regard to active monitoring, the InVS will send out follow-up questionnaires every three years to update the information on family status, health behaviour (e.g., smoking habits and participation in sports), place of residence, employment changes, and any changes in behaviour regarding personal protection equipment. These will be electronic questionnaires, to control costs (printing, mailing and data input costs) and to provide greater flexibility in content.

With regard to passive monitoring, the InVS plans to use the file from the Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM) [the national health insurance inter-plan information system] to monitor the emergence and frequency of health events (the protocol is currently being drawn up). This plan will ensure that the follow-up questionnaires sent to program participants are not too long, so as to limit the response burden. It will also reduce potential recall bias (inability to remember, difficulty situating events in time) or diagnostic bias (inadequate classification of disease) on the part of participants and guard against the risk of non-response (Kone Pefoyo, Rivard and Laurier, 2009).

The data provided by the SNIIRAM contain details about prescription drug claims and medical consultations, hospitalizations, work accidents and occupational diseases (Figure 2.3-1). Accessing these data entails, in particular, the use of the French national directory registration number (the NIR), a unique personal identifier. Under current legislation, use of the NIR outside the sphere of social protection requires a decree of the Council of State following the publication of an explanatory notice by the Commission nationale de l'informatique et des libertés (CNIL) [the national commission on information technology and liberties]. The InVS decided to go through a trusted third party that holds NIR information. The InVS must send the available identifying data (surname, maiden name, given names, sex, date of birth and place of birth) to this trusted third party by secure means. The third party then searches for the NIR and creates a file that matches the EpiNano program participation number with the information required to extract the SNIIRAM data (NIR, date of birth and sex). This file can then be transmitted by secure means to the health insurance authority, which is able to extract the corresponding health data or health insurance identifier and return it to the InVS (Figure 2.3-2). Under the protocol, information can be retrieved from the SNIIRAM for anyone who has not refused consent, that is, anyone who has not returned the non-consent form attached to the invitation message and the recruitment questionnaire. Consequently, this information will be available for some recruitment questionnaire non-respondents. Because of their nature and diversity, these data should help with the adjustment of collected data for partial non-response (missing data for recruitment questionnaire respondents) and total non-response (recruitment questionnaire non-respondents) (Santin et al. 2014).

Figure 2.3-1
Overview of information available in the Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM)

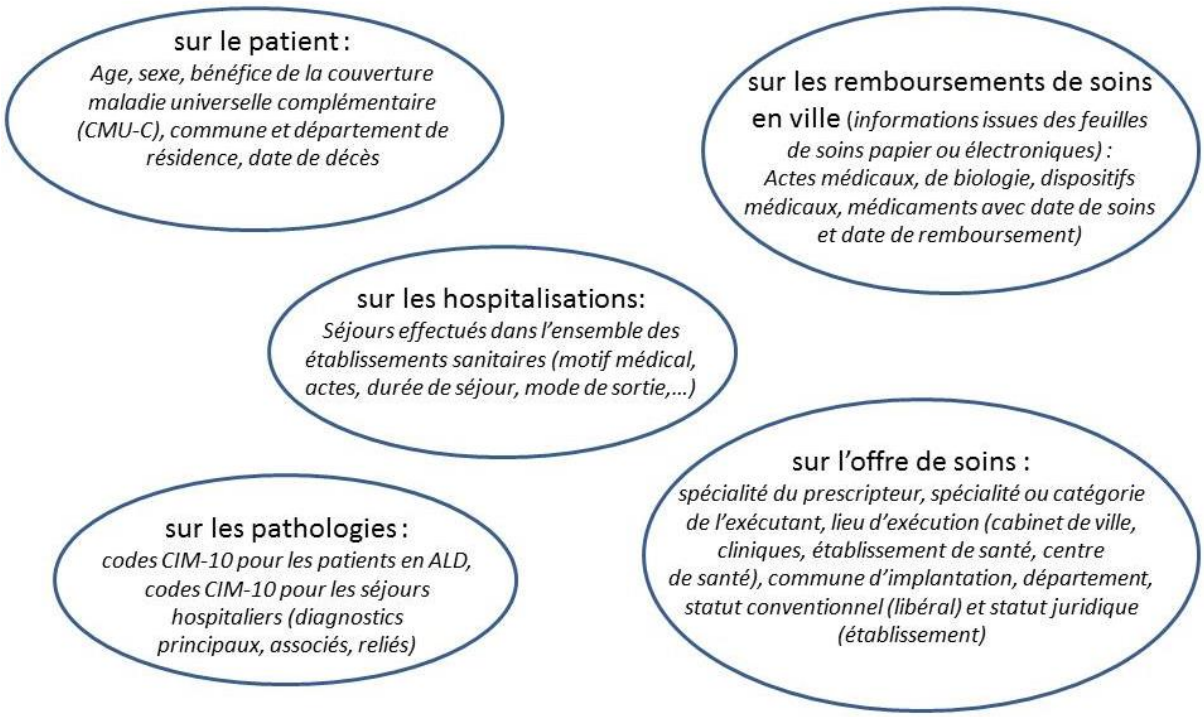
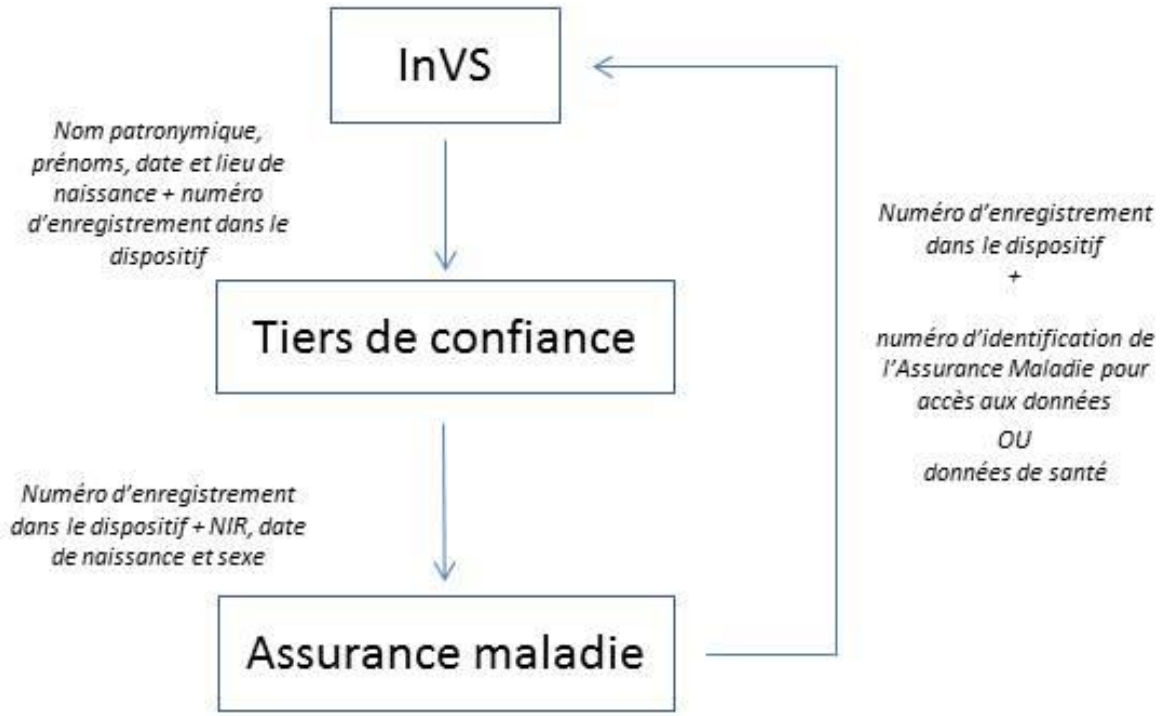


Figure 2.3-2
EpiNano program: Tentative collection diagram for data from the Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM)



In addition to describing the use of medical care by individuals, the data available in the SNIIRAM can be used to find pathologies with a drug- or procedure-selection algorithm, as many research projects have demonstrated (Moisan et al. 2011). Specific procedures or medications have been used to identify subjects who have certain diseases. For example, a prediction model for asthma was constructed on the basis of variables such as visits to a respirologist and claims for certain antiasthmatic drugs (Iwatsubo et al. 2012).

Through a search of the Répertoire national d'identification des personnes physiques (RNIPP) [the national directory for the identification of natural persons] and the database of medical causes of death (CépiDc), the InVS can collect the vital status of program participants and, where applicable, causes of death. To do so, the InVS has to send the surname (and maiden name), given names, sex, date of birth and place of birth of people who have not submitted non-consent forms to unit U1018 of the Centre de recherche en epidémiologie et santé des populations [the epidemiology and population health research centre], which is part of the Institut national de la santé et de la recherche médicale [the national institute of health and medical research]. This unit serves as a trusted third party in exchanges with the RNIPP and the CépiDc (CépiDc 2014).

3. Discussion

The InVS developed its surveillance of the occupational exposure of French workers to intentionally produced nanomaterials in response to a request by the directorate general of health and the directorate general of labour for more information about the health risks associated with nanomaterials. Under the research plan that was developed, the InVS will combine personal information from recruitment questionnaires with industrial hygiene data on the occupational exposure of participants and with information from medical–administrative databases, such as the number of visits to a family physician in the vital status record, prescription drug claims, and chronic-disease pathology reports. With this information, it is expected that the EpiNano program can meet its objectives of general monitoring of the potential medium- and long-term health effects of occupational exposure to nanomaterials.

There are uncertainties about the custodian of NIR information that will serve as a trusted third party in exchanges with the health insurance authority. The InVS has contacted the Caisse nationale d'assurance vieillesse [the national old-age insurance fund] in this regard, but there is no formal co-operation agreement between the two organizations.

The decision to use electronic questionnaires for monitoring is mainly a budgetary one. However, this may result in biases associated with the response mode. Not all participants necessarily have Internet access. In addition, the appearance and loading time of the questionnaire will vary depending on the computer technology used (e.g., desktop, laptop, touch-screen tablet, or smartphone). This may cause non-response biases that will have to be corrected later. This will have to be weighed against the cost savings.

The structure of the EpiNano program reflects an attempt to optimize data collection by fully exploiting the latest technologies. Other avenues remain to be explored, particularly with regard to participant retention through the development of a newsletter or a dedicated website, for example.

Following a planned three-year participation phase for businesses, the InVS plans to monitor participants over a period of twenty years, with, ultimately, an expected study population of 1,500 to 2,000 people. To date, the InVS has visited ten businesses, three of which have submitted the contact information of their employees. As a result, the InVS has sent out 54 questionnaires, with a return rate of 54%.

References

Boutou-Kempf, O. (2011). *Éléments de faisabilité pour un dispositif de surveillance épidémiologique des travailleurs exposés aux nanomatériaux intentionnellement produits*. Saint-Maurice: Institut de veille sanitaire.

Centre de recherche en épidémiologie et santé des populations. 2014. *Statut vital et causes de décès: Nouvelle procédure utilisateurs d'accès au RNIPP*. May 5, 2014. Available at: http://cesp.vjf.inserm.fr/~webifr/pdf/PROCEDURE_ACCES_RNIPP_UTILISATEURS_V2_19_05_2014.pdf.

European Union. 2011. *Recommendation of European Commission of 18 October 2011 on the definition of nanomaterial*.

Févotte, J., Luce, D., Arslan, M., Dananché, B., Delabre, L., Ducamp, S., Garras, L., Orlowski, E., Pilorget, C., Thuret, A., Goldberg, M., Imbernon, E. (2006). "Surveillance des expositions d'origine professionnelle en population générale : le programme Matgéné", *Bulletin épidemiologique hebdomadaire* 46–47: 362–365.

Canu, I.G., Boutou-Kempf, O., Delabre, L., Ducamp, S., Iwatsubo, Y., Marchand, J.L., *et al.* (2013). "French registry of workers handling engineered nanomaterials as an instrument of integrated system for surveillance and research", *Journal of Physics: Conference Series* 429 (1).

Guseva-Canu, I., Ducamp, S., Delabre, L., Audignon-Durand, S., Ducros, C., Durand, C., Iwatsubo, Y., Jezewski-Serra, D., Le Bihan, O., Malard, S., Radauceanu, A., Reynier, M., Ricaud, M. and Witschger, O. (2015). "Proposition d'une méthode de repérage des postes de travail potentiellement exposant aux nano-objets, leurs agrégats ou agglomérats dans les entreprises mettant en œuvre des nanomatériaux manufacturés", *Archives des maladies professionnelles*. Forthcoming.

Iwatsubo, Y., Lauzeille, D., Houot, M., Mevel, M., Chabault, E., Delabre, L. *et al.* (2012). *Surveillance épidémiologique de l'asthme au sein de la population des artisans et commerçants affiliés au Régime social des indépendants (RSI). Résultats de l'étude pilote*. Saint-Maurice: Institut de veille sanitaire.

Moisan F, Gourlet V, Mazurie JL, Dupupet JL, Houssinot J, Goldberg M, *et al.* (2011). "Prediction model of Parkinson's disease based on antiparkinsonian drug claims", *American Journal of Epidemiology*, 174 (3): 354–363.

Kone Pefoyo, A.J., Rivard, M., Laurier, C. (2009). "Public health surveillance and role of administrative data", *Revue d'épidémiologie et de santé publique* 57 (2): 99–111.

Pelucchi, C., Pira, E., Piolatto, G., Coggiola, M., Carta, P., and La Vecchia, C. (2006). "Occupational silica exposure and lung cancer risk: A review of epidemiological studies 1996-2005." *Annals of Oncology* 17: 1039–1050.

Santin, G., et al. (2014). "In an occupational health surveillance study, auxiliary data from administrative health and occupational databases effectively corrected for nonresponse" *Journal of Clinical Epidemiology* 67: 722–730.