

Development of a biomarkers database for the National Children's Study

Danelle T. Lobdell, Pauline Mendola*

*US Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory,
Human Studies Division, Epidemiology and Biomarkers Branch, MD 58A, Research Triangle Park, NC 27711, USA*

Received 17 March 2004; accepted 30 July 2004

Available online 9 April 2005

Abstract

The National Children's Study (NCS) is a federally-sponsored, longitudinal study of environmental influences on the health and development of children across the United States (www.nationalchildrensstudy.gov). Current plans are to study approximately 100,000 children and their families beginning before birth up to age 21 years. To explore potential biomarkers that could be important measurements in the NCS, we compiled the relevant scientific literature to identify both routine or standardized biological markers as well as new and emerging biological markers. Although the search criteria encouraged examination of factors that influence the breadth of child health and development, attention was primarily focused on exposure, susceptibility, and outcome biomarkers associated with four important child health outcomes: autism and neurobehavioral disorders, injury, cancer, and asthma. The Biomarkers Database was designed to allow users to: (1) search the biomarker records compiled by type of marker (susceptibility, exposure or effect), sampling media (e.g., blood, urine, etc.), and specific marker name; (2) search the citations file; and (3) read the abstract evaluations relative to our search criteria. A searchable, user-friendly database of over 2000 articles was created and is publicly available at: <http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=85844>. PubMed was the primary source of references with some additional searches of Toxline, NTIS, and other reference databases. Our initial focus was on review articles, beginning as early as 1996, supplemented with searches of the recent primary research literature from 2001 to 2003. We anticipate this database will have applicability for the NCS as well as other studies of children's environmental health.

Published by Elsevier Inc.

Keywords: Biomarkers; Children; Environmental health; Database

Introduction

Advances in biological markers (biomarkers) have facilitated the assessment of associations between environmental exposures and potential health effects by improving accuracy and precision of measures and helping to identify high-risk groups in exposed populations. Biomarkers can be measures of exposure, measures of susceptibility (showing either increased or decreased susceptibility to disease), or measures of effect, such as the presence of subclinical disease processes or diseases themselves. The measurement of all three types of biomarkers can be useful in the characterization of exposure–disease relationships.

The National Children's Study (NCS), a federally-sponsored longitudinal study currently beginning the implementation phase (Branum et al., 2003), has developed a biomarker's database to support the preparation of the study protocol. Other investigators interested in biomarkers for children's environmental health research can also use the database.

The NCS is expected to follow approximately 100,000 children from before birth until age 21 to examine the effects of environmental influences on health and development. More information about the NCS is available on the study website, www.nationalchildrensstudy.gov. Priority outcome areas of the NCS include pregnancy outcomes, child development (including developmental disorders), injury, asthma, and obesity (including altered growth and physical development).

* Corresponding author. Fax: +1 919 966 7584.

E-mail address: mendola.pauline@epa.gov (P. Mendola).

This project compiled relevant scientific literature to identify both routine or standardized biological markers as well as new and emerging biological markers that would be relevant to the NCS.

Methods

A biomarkers database was created in 2002, which compiled relevant scientific literature, mostly review papers, on biomarkers. A subsequent update was performed in 2003. The literature search was primarily conducted in the PubMed (National Center for Biotechnology Information) search engine with some supplementary searches from TOXLINE (National Library of Medicine, 2004a, 2004b, 2004c) and National Technical Information Service (National Technical Information Service, 2004).

A keyword search was based on a preconceived definition of biomarkers (see Table 1) and four principal child health outcomes (although the search criteria did encourage examination of other factors that influence the

breadth of child health and development)—cancer, asthma and respiratory health, autism and neurodevelopment, and injury. Table 2 presents the primary keyword search strings with limitations.

Each search return was displayed in “MEDLINE” format (National Library of Medicine, 2003), then sent to file and saved as a *.txt file. In cases where a return was too large to allow PubMed to export a file directly (greater than 500 records), the return was displayed as text in the browser and saved from the browser. Search returns containing no references were not saved.

Citations from the set of text files were sorted and managed using Reference Manager (RefMan), a proprietary product of ISI ResearchSoft. An original database was created in RefMan and the *.txt files were imported using the “Import Text File” function under a PubMed.cap filter. Preferences were set such that duplicate citations were not imported. RefMan supplied each reference with a non-duplicate, non-sequential reference identification number upon import. These numbers were used during citation review to eliminate duplication of effort among the review team members.

Each abstract was reviewed and scored to identify those with relevant biomarker information. A “Pass/Fail Criteria for Abstract Review” document was created to score the abstracts (see Appendix A). This document was also used to reduce inter-rater variability among the review team. The team read each title and abstract (when abstract was included) and assigned pass/fail/cannot ascertain.

A Biomarker Information Form was completed for each passed abstract. The review form collected information such as biomarker name, type, health outcome category, specific health outcome, and specimen type. Each passed abstract yielded one or more biomarker names and the associated health outcome. The completed forms were incorporated into the Biomarkers Database.

Failed abstracts were given a ‘reason for failure’ that included “no health outcome”, “prognosis only”, “treatment only”, “medical test”, “clinical trial”, “research methods”, “biomarker measured in tumor only”, “other”, and “irrelevant”. The “irrelevant” category included abstracts discussing the elucidation of a biological pathway or signal cascade with no mention of a specific biomarker or adverse effect. The “other” category included citations with no adverse effect and, when obvious from the title, all of the preceding categories.

Once each abstract was assigned with a value of pass/fail/cannot ascertain, the data were loaded directly from RefMan into Microsoft Access 2000 Database.

Results

The Biomarkers Database contains all references obtained during the 2002 and subsequent 2003 update literature search, including those abstracts marked as Cannot Ascertain or Fail. The database includes 4736 citations of which 1398 citations satisfied the screening criteria. Most citations are

Table 1
Definitions of biomarkers and biomarker categories

Biomarker

A biomarker is defined as a cellular, biochemical, or molecular (or tissue) alteration that is measurable in biological media, such as human tissues, cells, or fluids; or the product of an interaction between a xenobiotic agent and some target molecule. Biomarkers include specific genes, pre-cancerous lesions or polyps, and diagnostic markers. For this task, biomarkers do *not* include:

- markers of prognosis only;
- medical or clinical test results, such as from routine medical blood testing (blood lipids, white blood cell counts, sepsis workups, bone mineral density, EEGs, or x-ray analysis);
- behavioral/cognitive functioning test results;
- growth measurements, such as birth weight or length, or head circumference;
- other physical measurements or observations (such as spirometry results, heart rate, blood pressure, or fingerprint ridge count); or
- mutagenicity tests.

Biomarkers of exposure

Biomarkers of exposure include concentrations of the exogenous parent chemical, its metabolites, or reaction products in body fluids or tissues, for example, blood lead. *Bioindicators* reflect only the presence or absence of a substance. *Biomonitorors* provide for regular surveillance and quantification of the amount of substance present.

Biomarkers of effect

Biomarkers of effect are measures of disease progression (e.g., biomarkers of early loss of pregnancy, benzo-pyrene-DNA adducts, antigen production, tumor secretions, gene suppression, etc.). Biomarkers measured in tumor tissue were *not* included, as the disease is diagnosed prior to the biomeasure, and the biomarker is used to determine prognosis rather than effect.

Biomarkers of susceptibility

Biomarkers of susceptibility include the detection of a particular genotype or polymorphism (such as genetic markers of cancer susceptibility), or may be a precursor disease, such as pre-cancerous colonic polyps. They may indicate the presence of or potential for sub-clinical disease or may indicate a potential protection against negative health effects of the exposure.

Table 2
Summary of keyword search strings and search limitations

Category	Keyword search strings	Limitations
General	1. Biomarker and exposure and environmental pollution	1999–7/17/03; Review; English
	2. Molecular epidemiology	
	3. Biomonitor and exposure	
	4. Genetic marker and biomarker	
	5. Predisposition to disease and biomarker	
	6. Metabolite and exposure	
	7. Fetus and biomarker	
	8. Pregnancy complication and biomarker	
Exposure	9. Pesticides and biomarker	1999–7/17/03; Review; English, 10/01/02 to 7/25/03; Primary articles
	10. Air pollution and biomarker	
	11. Effect biomarker and exposure	
	12. Susceptibility biomarker and exposure	
Asthma	13. Asthma and biomarker	1999–7/17/03; Review, English; 10/01/02 to 7/25/03; Primary articles
	14. Respiratory health and biomarker	
	15. Asthma and biomarker and exposure	
	16. Respiratory tract diseases and biomarker	
Cancer	17. Cancer and biomarker	2001–7/17/03; Review, English
	18. Cancer and biomarker and pregnancy	
	19. Cancer and biomarker and pregnancy and exposure	
Neurodevelopmental Disorders	20. Autism and biomarker	1996–7/17/03; Review, English, 10/01/02 to 7/17/03; Primary articles
	21. Neurodevelopmental disorder and exposure	
	22. Developmental disability and biomarker	1999–7/17/03; Review, English
	23. Neurotoxic* and biomarker	
	24. Neurological development and biomarker	

Table 2 (continued)

Category	Keyword search strings	Limitations
Injury	25. Injury and biomarker	1996–7/17/03; Review, English
	26. Injury and exposure and adverse effects	
	27. Coordination and biomarker and exposure	1999–7/17/03; Review, English
	28. Impulsive behavior and biomarker and exposure	
Miscellaneous		2000–7/17/03; Review or representative reference

review articles, beginning as early as 1996, supplemented with searches of the recent primary research literature from 2001 to 2003 (Table 2). A non-editable version of the database, installation instructions and a detailed users guide is publicly available at <http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=85844>.

The Biomarkers Database was designed to serve three purposes: (1) to allow users to search the biomarker records compiled, (2) to allow users to search the citations file, and (3) to allow users to read the abstract reviews. Most of the biomarker information recorded in the database was obtained from abstracts. During the abstract review process, each reviewer recorded in the Biomarker Information form the name of any biomarker(s) that were discussed in the reference's abstract. In addition, if available, the biomarker type, health outcome, specimen type, and exposure were recorded.

The biomarker records search procedure allows researchers to search the Biomarkers table in the database by defining a search term for one or more fields. The user can specify search terms for any or all of the following fields: Biomarker Name, Health Outcome Category, Health Outcome Description, Biomarker Type, Exposure, or Specimen Type. The user can specify whether the search is to be conducted only on abstracts that passed the review process (which is the default) or on all abstracts in the database.

The citations search procedure is conducted on the citations compiled from the literature search. The citations search procedure will allow the researcher to search multiple fields by defining a search term for each field. The user can specify search terms for any of the following fields: Title, Author, Journal, Year of Publication, or Keyword. Similar to the biomarkers records search procedure, the user can specify whether the citation search is to be conducted only on passed abstracts or on all abstracts in the database.

The third main feature of the database is the ability to read notes from a review of the article and/or its abstract. These review forms are similar to those used by project staff during abstract and article review.

There are three review forms:

- The Abstract Review form is used to record whether the article may be relevant to this task. Every abstract in the biomarker database will have a pass, fail, or not ascertained result code on this form.
- The Biomarker Information form is used to record any biomarkers referenced in the abstract along with its health outcomes and some other information.
- The Additional Information (Article Review) form was designed to record information obtained from reading the actual article.

A report can be created from the various searches for: (a) all titles in the list, (b) a highlighted title, or (c) multiple highlighted titles. The various reports include:

- a printout of citations only;
- a printout of citations with biomarker information;
- a printout of citations with abstract.

Discussion

Human exposure assessment typically involves some level of (1) measurement of the concentration of the contaminant in the appropriate media (e.g., air, water, food); (2) estimation of exposure duration with respect to each medium; and (3) calculation of the approximate total absorption from all media. Building on that information, biomarkers of exposure (i.e., concentrations of the parent chemical, its metabolites, or reaction products in body fluids or tissues) integrate the effects of absorption, metabolism, disposition, and elimination of the chemical by the human body. Thus, biomarkers (either *bioindicators* which reflect only the presence or absence of a substance or *biomonitors* which provide for regular surveillance and quantification of the amount of substance present) can provide a relatively direct measure of tissue dose, which may be used as the basis for more accurate exposure assessments. Established techniques and those that are currently under development can help improve the validity and reliability of body burden measures of a wide spectrum of environmental exposures for the NCS and other studies of children's environmental health.

Ideally, an exposure biomarker measurement should have a known relationship to the exposure of interest as well as the biologically effective internal exposure. To be successfully implemented in population-based studies, biomarkers are most valuable if they lend themselves to routine monitoring with respect to acceptability (will people agree to the sampling) and cost-effectiveness as well as quantification.

Biomarkers of susceptibility and effect are equally relevant to the evaluation of the health and safety risks associated with exposure to toxic pollutants and chemicals. Biomarkers of susceptibility (e.g., the detection of a particular genotype or polymorphism) can indicate the presence of or

potential for subclinical disease or may indicate a potential protection against negative health effects of the exposure. Such biomarkers may be considered either intermediate outcomes or covariates acting as effect modifiers in the exposure–disease relationship. Biomarkers of effect are considered measures of disease progression. Longitudinal studies, such as the NCS, provide the ideal setting for identifying such intermediate endpoints and markers of subclinical disease since study participants enroll before the outcomes of interest occur.

We anticipate updating the publicly available database approximately annually as the NCS progresses. Additional search terms, such as terms related to biomarkers of inflammation, may be added to the database in the future as additional markers of interest for the NCS are explored.

Acknowledgments

The authors would like to acknowledge the contributions of Westat, Inc. (EPA 68-D-98-115) and RTI International (EPA 68-D-02-069) in the development of the database and review of abstracts for inclusion.

Appendix A. Pass/Fail criteria for abstract review

Step 1: To pass the abstract review, the abstract must:

- 1) Describe research conducted on humans or animals during any life stage.
- 2) Describe one or more biomarkers, as defined (see Table 1).
- 3) Describe any of the topics present in the YES column but none in the NO column of the table below.

YES	NO
–Investigation of specific genes	–markers of prognosis only
–pre-cancerous lesions or polyps	–medical or clinical test results
–diagnostic markers	–behavioral/cognitive
–exogenous parent chemical in body fluids or tissues	functioning test results
–metabolites in body fluids or tissues	–physical measurements such as birthweight
–reaction products in body fluids or tissues	–biomarkers measured in tumor tissue
–antigen secretion	
–gene suppression	
–DNA adducts	
–susceptible genotype	
–susceptible polymorphism	
–precursor disease	

- 4) Describe an adverse health outcome.

An abstract containing insufficient information to answer the above 4 conditions were marked as "cannot ascertain".

Step 2: Each abstract requires the entry of pass/fail/cannot ascertain into the Abstract Review Form of the Biomarkers Database. If the abstract meets all of the criteria, a Biomarker Information Form is completed.

References

- Branum, A.M., Collman, G.W., Correa, A., Keim, S.A., Kessel, W., Kimmel, C.A., Klebanoff, M.A., Longnecker, M.P., Mendola, P., Rigas, M., Selevan, S.G., Scheidt, P.C., Schoendorf, K., Smith-Khuri, E., Yeargin-Allsopp, M., The National Children's Study Interagency Coordinating Committee, 2003. The national children's study of environmental effects on child health and development. *Environ. Health Perspect.* 111, 640–646.
- National Library of Medicine (Accessed February 12, 2004). National Center for Biotechnology Information. PubMed. Available at: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>.
- National Library of Medicine (Accessed February 12, 2004). U.S. National Library of Medicine MEDLINE Distributed in XML Format. Available at: http://www.nlm.nih.gov/bsd/licensee/data_elements_doc.html.
- National Library of Medicine (Accessed February 12, 2004). TOXLINE. Available at: <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?TOXLINE>.
- National Technical Information Service (Accessed February 12, 2004). Available at: <http://www.ntis.gov/>.