Transportation, Space, and Technology

A RAND INFRASTRUCTURE, SAFETY, AND ENVIRONMENT PROGRAM

This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

Jump down to document ▼

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

Purchase this document

Browse Books & Publications

Make a charitable contribution

For More Information

Visit RAND at www.rand.org

Explore the RAND Transportation, Space, and Technology Program

View document details

Limited Electronic Distribution Rights

This product is part of the RAND Corporation documented briefing series. RAND documented briefings are based on research briefed to a client, sponsor, or targeted audience and provide additional information on a specific topic. Although documented briefings have been peer reviewed, they are not expected to be comprehensive and may present preliminary findings.

Effects of Preanalytical Variables on the Quality of Biospecimens Used to Study Genetic Changes in Cancer

Development of the Biospecimen Research Database

Elisa Eiseman

RAND

Transportation, Space, and Technology

A RAND INFRASTRUCTURE, SAFETY, AND ENVIRONMENT PROGRAM

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

SUMMARY

Human biospecimens[1] are valuable research tools because they reflect the state of the biospecimen at the time it was collected. That is, the expression pattern of genes and proteins depends on both the biological state of the biospecimen (e.g., whether it is lung or colon tissue; whether it is diseased or normal) and the environmental and biological stresses the biospecimen experiences prior to analysis (i.e., preanalytical variables). Examples of preanalytical variables include

- medical or surgical procedures conducted before and during the removal of the biospecimen from the patient (e.g., administration of antibiotics, anesthesia, and other drugs; disruption of blood supply to the tissue; or intraoperative administration of blood, blood products or other fluids)

- biospecimen-processing methods (e.g., type of fixative, time in fixative, method and rate of freezing)

- duration and conditions of biospecimen transport and storage (e.g., storage and transport temperature, duration of storage).

Molecular analyses of biospecimens from cancers and other diseases have revealed changes in gene and protein expression (e.g., either over- or underexpression of specific genes). It is interesting to note that many of the same genes reported to have altered expression in diseases have also been shown to change expression in response to environmental changes and biological stresses. For example, it is clear from studies on yeast, plants, and animals that changes in temperature, pH, and nutrient availability; oxygen deprivation; and other environmental stresses can cause major changes in gene expression (Storey and Storey, 2001; Steinberg, Stürzenbaum, and Menzel, 2008; Kenneth and Rocha, 2008; Van Elzen, Moens, and Dewilde, 2008). Significant changes in gene expression can occur as early as 15 minutes after exposure to a stimulus or stress, while posttranslational changes in proteins, such as methylation and

---

[1] Human biospecimens include everything from subcellular structures (e.g., DNA, mRNA, proteins) to cells, tissue (bone, muscle, connective tissue, and skin), organs (e.g., liver, bladder, heart, kidney), blood, gametes (sperm and ova), embryos, fetal tissue, and waste (urine, feces, sweat, hair and nail clippings, shed epithelial cells, placenta).

phosphorylation, can occur within seconds (Eastmond and Nelson, 2006; Kawasaki et al., 2001; Eiseman and Bolen, 1992). Since the value of a biospecimen to a researcher is the information it contains about the actual biological state of the specimen as it existed in the person from whom it was derived, determining which changes are disease-related and which are artifacts caused by preanalytical variables is of utmost importance.

The scientific community has repeatedly identified the limited availability of carefully collected and controlled, high-quality human biospecimens annotated with essential clinical data and properly consented for broad investigational use as the leading obstacle to progress in postgenomics cancer research (OBBR, undated [c]). The National Cancer Institute (NCI) is leading a national initiative to systematically address and resolve this problem. Since 2002, when NCI leadership identified biorepositories as an area of critical importance, NCI has been involved in several efforts to determine best practices for biospecimen collection and management. In support of this effort, NCI established the Office of Biorepositories and Biospecimen Research (OBBR) in 2005 to address the issues associated with the need for high-quality, well-annotated biospecimens for biomedical research.

OBBR's mission is "to ensure that human specimens available for cancer research are of the highest quality" (OBBR, undated [a]). To accomplish its mission, OBBR has established biobanking as a new area of research and conducts and funds research on the effects of preanalytical variables on the usefulness of biospecimens in genomic and proteomic studies (OBBR, undated [c]). The results of the research sponsored by OBBR will support the development of guidelines and evidence-based standards for biospecimens and biorepositories that will optimize the quality and accessibility of biospecimens for the cancer and broader biomedical research communities.

One of the questions in which OBBR was interested was what, if any, data exist on the effects of preanalytical variables on biospecimens. To begin to answer this question, OBBR asked RAND to identify and analyze existing data on the effects of these variables on biospecimens used to study genetic and proteomic changes in cancer. The full implementation of this project was envisioned as a multiyear project consisting of three objectives: (1) to identify and analyze existing data on the effects of preanalytical variable on biospecimens used to study genetic and proteomic changes in cancer; (2) to create an interactive, searchable Web site that scientists, pathologists, repositories, and others can visit to learn

about and contribute data, methods, and other relevant information on how biospecimens used to study genetic and proteomic changes in cancer are affected by preanalytical variables; and (3) to provide information to the research community and other interested parties about the effects of preanalytical variables on the quality of biospecimens used to study genetic and proteomic changes in cancer. The information generated by this project was intended to provide OBBR with insight into the molecular impacts of different preanalytical variables on different biospecimen types, research questions, and analysis methods.

This documented briefing, which focuses on work conducted during the first year of this multiphase project, describes the process used to identify and analyze data on the effects of preanalytical variables on biospecimens used to study genetic and proteomic changes in cancer. It provides details on the development of the Biospecimen Research Database, a data-curation tool developed to provide a standardized way of consistently recording data on the effects of preanalytical variables, and summarizes the findings of the first phase of the study.

Developing the Data-Curation Tool

To make the findings of this project useful to the scientific community, it was necessary to develop a systematic way of capturing the wealth of data collected through the review of the scientific literature. A data-curation tool, called the Biospecimen Research Database, was developed to provide a standardized way of consistently recording data obtained through the literature review. Developing the data-curation tool involved several activities. First, the major subject-area headings and specific fields for data collection had to be defined. Next, a data-accession tool needed to be designed. A preliminary template was designed using a Microsoft Excel spreadsheet, which was pilot tested to determine whether the appropriate data-collection fields had been selected. A more user-friendly, interactive data-accession tool was then designed using a Microsoft Access database, which was also pilot tested to assess its usability and robustness.

The first step in developing the data-curation tool was to determine the types of data that would be collected from the literature review. The types of data to be collected were grouped into six major subject areas of interest: biospecimen type, tissue type, diagnosis, biomolecule type, technology platform, and experimental factors. Next, specific fields for data collection within each major subject-area heading were identified. The major subject-area headings and associated specific data-entry fields

went through several revisions during the development of the Excel and Access data-curation tools.

A preliminary data-collection template was developed using an Excel spreadsheet. The Excel data-collection template was pilot tested to assess its usability and robustness. The template was refined by the addition and deletion of data-entry fields based on the type and importance of information found during the pilot test.

The Access database curation tool was developed in collaboration with OBBR. The fields defined in the Excel data-collection template formed the basis of the Access database curation tool. Data-entry forms with drop-down menus and free-text boxes were developed to improve the ease of data entry. Two different forms were developed: one to capture general information about the paper and one to capture specific data about the studies within the paper. The Access database curation tool was pilot tested to provide a direct comparison to the Excel template, confirm its usability and robustness, and provide indications of where revisions of data-entry fields were necessary.

The Access database curation tool formed the basis for the development of an online data-collection Web site. The online data-collection Web site, which is still under development, was designed in such a way that, as data are entered, they directly populate an online, searchable database. The data-collection Web site and the searchable Biospecimen Research Database, which contains data on the effects of preanalytical variables on the quality of biospecimens, is being built by OBBR with input from RAND and hosted on the OBBR external Web site. A prototype version is available online as a Web-based, searchable database that provides information about the effects of preanalytical variables on biospecimens (see OBBR, undated [d]).

Literature-Search Strategies

A comprehensive search of the scientific literature was performed to identify studies conducted specifically to determine the effects of preanalytical variables on the quality of biospecimens used to study genetic and proteomic changes in cancer. To accomplish this, RAND developed a literature-search strategy designed to find studies of interest. Papers were selected to populate the database using several literature-search strategies, including keyword searches, MeSH® term searches, and author searches.

Keyword and MeSH term searches of PubMed were conducted using a targeted set of search terms to find relevant articles. Searches ranged from very specific (such as the keyword search for *preanalytical variable* and variations thereof and the MeSH term search for *tissue fixation*) to very broad, general searches (such as those using variations of the terms *human*, *specimen*, *acquisition*, *processing*, *storage*, *effect*, *gene*, *protein*, *DNA*, *RNA*, and *analysis*). Other relevant Web sites were also searched, including journals that feature biological methods (e.g., *BioTechniques*, *Cell Preservation Technology* [now *Biopreservation and Biobanking*]). Searches for relevant studies also included the examination of the reference lists of articles already retrieved. In addition, a search of PubMed was performed to identify relevant papers authored by a target list of investigators who are active in the field of biospecimen research. Of all the searches, the search using the MeSH term *tissue fixation* yielded the highest percentage of relevant papers. Virtually all of the papers identified by this search were relevant, and almost 30 percent of the papers were analyzed and included in the database.

The time period specified for the searches covered the past 20 years (i.e., from 1987 through 2007). While many relevant studies were found in papers published more than 10 years ago (i.e., papers published before 1997), it was decided that more recent papers (i.e., papers published between 1997 and 2007) would be of most relevance to the research community and were selected as a place to start to populate the database. Search results were also limited to English-language publications. Only studies that used human biospecimens were included in the database; studies using biospecimens from other animal sources were not analyzed. Also, only original research articles were included in the database; review articles were not included.

Data Analysis and Study Results

The effects of preanalytical variables on the molecular profile of biospecimens will differ depending on the specimen type (e.g., blood, urine, normal tissue, cancerous tissue), the biomolecule being analyzed (e.g., DNA, RNA, protein), the analysis method being used (e.g., Southern blot, polymerase chain reaction [PCR], fluorescent in situ hybridization [FISH], cDNA microarrays), and the research question being asked. One way to conceptualize these preanalytical variables and their potential effects is to array them according to the biospecimen type, the analysis method, and the research question being asked (Figure S.1) (Barker et al., 2005). The three-dimensional array has the biospecimen types along the x-axis, the biomolecule types and associated analysis methods along the y-

axis, and the research questions along the z-axis (the blue, red, and off-white colors depict different research questions). This array, developed by OBBR, provides a useful framework for the analysis of the effects of preanalytical variables on biospecimens. By systematically filling in the boxes in the array for each biospecimen type with information about the effects of preanalytical variables on the technology platform used and the research question asked, insight can be gained into the specific impact of different preanalytical variables on the molecular profile of the biospecimen.

At the time of the briefing to OBBR, 145 studies from 65 papers had been analyzed and entered into the database. The number of studies per paper ranged from one to eight, with most papers containing two to three studies. Of the 145 studies, 45 studies analyzed DNA as the biomolecule,

Figure S.1. Framework for Analysis of Effects of Preanalytical Variables on Biospecimens
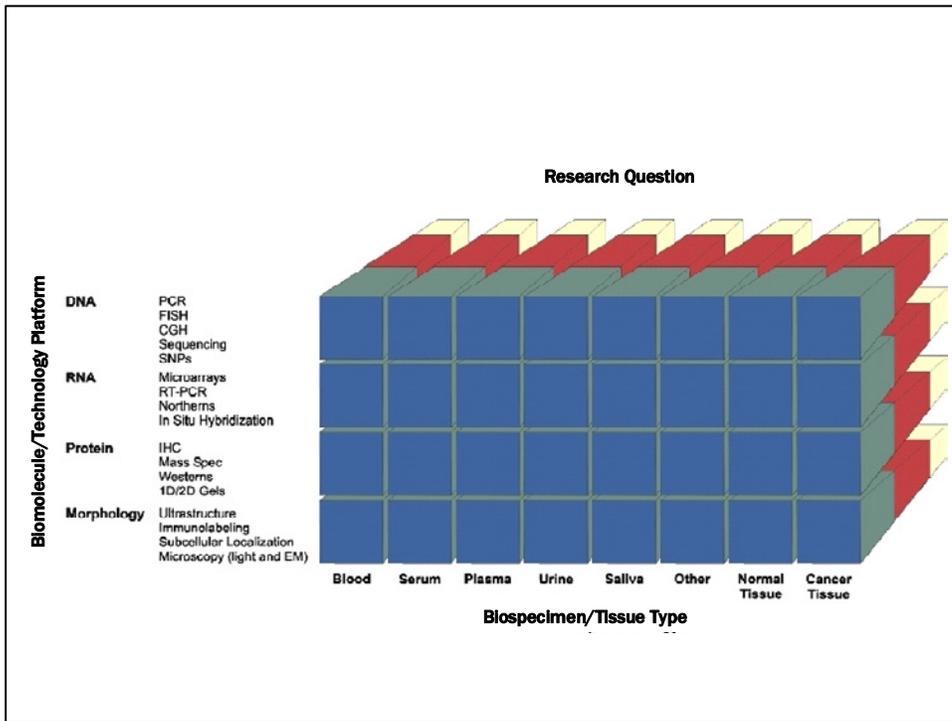
46 analyzed RNA, 53 analyzed proteins, and 10 analyzed morphology[2] (note that each study may include more than one biomolecule type). Currently, there are data from 193 studies from 80 papers in the database.

The studies used 22 different technology platforms and reported on 15 preanalytical variables. The most commonly used technology platforms were reverse transcription PCR (RT-PCR) to analyze RNA, immunohistochemistry to analyze proteins, and PCR to analyze DNA. The most commonly investigated preanalytical variables were type of fixative, biomolecule extraction method, time at room temperature/pre-fixation time, and time in fixative. Most preanalytical variables had several associated values. The number of associated values ranged from one to eight, with most preanalytical variables having two associated values. For example, *time in fixative* typically had several associated values, since each time point in an experiment using a time course would be a new value (e.g., 15 minutes, 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 5 hours).

Challenges and Next Steps

There are some challenges when it comes to analyzing the data in the Biospecimen Research Database. Data on the effects of preanalytical variables on biospecimens are not reported consistently in the literature, a fact that may make comparisons between studies and analyses across studies difficult (i.e., meta-analysis). For example, one study may report time in weeks (e.g., 4 weeks), while another may report it in months (e.g., 0 to 1 month). It is important to determine whether these different measures of time are comparable when performing meta-analyses. The database may be most useful as a tool to identify gaps in research on the effects of preanalytical variables on biospecimens in which additional studies may be valuable.

Consistency in recording data in the database is also crucial to be able to make comparisons between studies and perform meta-analyses. Drop-down menus with controlled vocabularies were used to help prevent

---

[2] *Morphology* is the study of the size, shape, and structure of a particular organism, organ, tissue, or cell. For the purposes of this briefing, the term *morphology* refers to the examination of the detailed structure of cells and tissues (sometimes called *cellular morphology*). Techniques used to study the morphology of cells and tissues include histology, immunohistochemistry, in situ hybridization, electron microscopy, and light, fluorescent, and confocal microscopy.

variation from being introduced into the database by the curation process. Limiting or even eliminating the use of free-text boxes to record important findings would also be helpful. Another way in which variation was controlled was by using a second reviewer to check the accuracy of the entries in the database and ensure consistency across data entered by different curators.

Another challenge in analyzing the data is the way in which the effects of preanalytical variables (i.e., the results of the studies) are recorded. Currently, preanalytical variables are selected from a drop-down menu in the data-curation tool, allowing researchers using the Biospecimen Research Database to easily identify which preanalytical variables were investigated in the study. In contrast, the results of the study are recorded in a free-text box (i.e., "Summary of Findings"), making it more difficult to identify what effect, if any, the preanalytical variable had on the biospecimen used, the biomolecule analyzed, or the research question asked. A more systematic way is needed to record and easily identify which preanalytical variables had an effect and what those effects were.

The next steps for the Biospecimen Research Database include expanding the information in the database with data from additional studies that focus directly on the effects of preanalytical variables on biospecimens, as well as adding information from clinical laboratory testing procedures relevant to research on genetic and proteomic changes in cancer (e.g., genetic testing, cytogenetics, molecular pathology). Information may also be obtained from studies that address preanalytical effects as part of the methodology section of the paper. In addition, information may be available from technical-support documents that accompany products used to collect, process, store, transport, or analyze biospecimens (e.g., DNA and RNA purification kits, DNA sequencers, real-time PCR machines). Eventually, it may be feasible to obtain unpublished data on the effects of preanalytical variables on biospecimens from investigators who are active in the field of biospecimen research.

As the database grows, it will be possible to fill in more boxes in the array and to fill each box with sufficient information to be able to start performing analyses of the effects of preanalytical variables on the biospecimens. These data could be used to identify gaps in knowledge about the effects of preanalytical variables and to support the development of guidelines and evidence-based standards for the collection, processing, and storage of biospecimens. The ultimate goal of the database is to provide information to OBBR and the scientific

community that will optimize the quality, accessibility, and utility of biospecimens for research purposes.