

A Call to Standardize Preanalytic Data Elements for Biospecimens, Part II

James A. Robb, MD; Lynn Bry, MD, PhD; Patrick M. Sluss, PhD; Elizabeth A. Wagar, MD; Mary F. Kennedy, CT(ASCP), MPH; for the College of American Pathologists Diagnostic Intelligence and Health Information Technology Biorepository Working Group 2

• **Context.**—Biospecimens must have appropriate clinical annotation (data) to ensure optimal quality for both patient care and research. Additional clinical preanalytic variables are the focus of this continuing study.

Objective.—To complete the identification of the essential preanalytic variables (data fields) that can, and in some instances should, be attached to every collected biospecimen by adding the additional specific variables for clinical chemistry and microbiology to our original 170 variables.

Design.—The College of American Pathologists Diagnostic Intelligence and Health Information Technology Committee sponsored a second Biorepository Working Group to complete the list of preanalytic variables for annotating biospecimens. Members of the second Biorepository Working Group are experts in clinical pathology and microbiology. Additional preanalytic area-specific variables were identified and ranked along with definitions

The College of American Pathologists (CAP) is committed to helping pathologists and the medical specialty of pathology successfully implement genomically informed, personalized precision health care. As part of this mission, the CAP published a practical resource for community pathologists.¹ This publication promotes the creation of high-quality biospecimens and biorepositories and supports the emerging opportunities in the biorepository field. Definitions of a biospecimen and a biorepository are given in that article. The subsequent publication provided 170 preanalytic variables that can, and in some instances should, be used for the creation of high-quality biospecimens that

and potential negative impacts if the variable is not attached to the biospecimen. The draft manuscript was reviewed by additional national and international stakeholders.

Results.—Four additional required preanalytic variables were identified specifically for clinical chemistry and microbiology biospecimens that can be used as a guide for site-specific implementation into patient care and research biorepository processes.

Conclusions.—In our collective experience, selecting which of the many preanalytic variables to attach to any specific set of biospecimens used for patient care and/or research is often difficult. The additional ranked list should be of practical benefit when selecting preanalytic variables for a given biospecimen collection.

(*Arch Pathol Lab Med.* 2015;139:1125–1128; doi: 10.5858/arpa.2014-0572-CP)

are common to all biospecimens or are specific to anatomic pathology, blood, body fluids, and nucleic acids.² Preanalytic variables specifically for clinical chemistry and microbiology were not within the scope of that study and are the subject of the present study.

A critical ingredient in the process of creating high-quality and high-value biospecimens is the clinical and processing information attached (annotated) and accrued to each biospecimen. Currently, only limited and/or relatively general recommendations are available as to what preanalytic information should be collected and attached to each biospecimen for clinical care and research. This information was identified and discussed in the previous publication.² Furthermore, there is great variability in the preanalytic information that is currently collected, as well as variability in the quality of this information. A dynamic list of evidence-based and/or expert opinion-defined preanalytic, clinical data fields is very useful for biospecimen annotation in the scope of both patient care and research. An additional benefit of this list should be the future usefulness of being able to retrospectively reassess or filter the overall quality of the biospecimen based on well-annotated data. This, in turn, will permit higher-quality clinical testing and research as new discoveries emerge about how key variables impact the quality of biospecimens.

While not every biorepository is embedded in, or derived from, a clinical laboratory setting, every organization that collects patient biospecimens must be the custodian for

Accepted for publication December 4, 2014.

Published as an Early Online Release January 16, 2015.

From the Department of Pathology, Brigham and Women's Hospital (Dr Bry), and the Department of Pathology, Massachusetts General Hospital (Dr Sluss), Boston; the Department of Pathology, MD Anderson Cancer Center, Houston, Texas (Dr Wagar); and the College of American Pathologists, Northfield, Illinois (Ms Kennedy). Dr Robb is a consulting pathologist in Boca Raton, Florida.

The authors have no relevant financial interest in the products or companies described in this article.

The opinions expressed represent those of the authors and are not to be construed as official College of American Pathologists guidelines or recommendations.

Reprints: Mary Kennedy, CT(ASCP), MPH, College of American Pathologists, 325 Waukegan Rd, Northfield, IL 60093-2750 (e-mail: mkenned@cap.org).

Table 1. Clinical Chemistry and Microbiology Biospecimens

Variable (Data Field)	PR ^a	Definition/Note	Negative Impact If Data Missing ^b
Fasting: length of fasting period (C)	1.0	How long has the patient been fasting?: time in hours (eg, 6.5 h or a range of hours)	1, 5, 8, 9, 10
Fasting: measured or observed (C)	1.0	Was the length of the fasting period measured and reported by the patient, or was it observed by staff?	1, 4
Transport medium (M)	1.0	For example, type of viral transport medium, anaerobic container, open to air with or without ice, stool fixed for ova and parasite evaluation	1, 6, 8, 10
Period between biospecimen collection and processing (M)	1.0	For example, biospecimens collected for RNA virus evaluation need to be processed within 4 h of collection (eg, human immunodeficiency virus)	1, 2, 4, 8, 10

Abbreviations: C, clinical chemistry; M, microbiology; PR, priority number.

^a Priority number 1.0 is a required field.

^b Negative impact definitions are linked to the numbers in Table 2. These impacts are examples and may not represent all the negative impacts for the variable.

these biospecimens, whether the custodianship is short term or long term. Examples of the periods associated with custodianship in a clinical setting before the samples are used (or reused) in clinical diagnostic testing include the following: hours to days for biospecimens that are being transferred to another site for testing and/or storage; 1 to 2 weeks for clinical pathology biospecimens such as blood, body fluids, or microbiology material; weeks to years for biospecimen-derived products such as DNA and RNA; and years for formalin-fixed, paraffin-embedded tissues and their associated glass and digital slides. This custodianship is crucial for the creation and maintenance of high-quality clinical biospecimens suitable for research. Biospecimens collected today may need to be used for clinical care and research in the future. Recapturing preanalytic data elements not recorded at the time of collection would prove difficult, if not impossible.

MATERIALS AND METHODS

The CAP Diagnostic Intelligence and Health Information Technology Committee's Biorepository Working Group 2 members were selected for their domain expertise in the areas of clinical chemistry and microbiology. Their input was voluntary, and no conflicts of interest were identified specific to the topic of this activity.

We defined preanalytic variables as those variables that may affect the quality and/or value of a biospecimen from the time of consenting until the biospecimen is used for testing either in direct patient care or in research.² The quality of a biospecimen is determined by its suitability for the purposes for which it was collected. Currently, messenger RNA, phosphorylated proteins, and peptides require the most stringent quality parameters because the technologies used to evaluate these targets are finicky (not robust) with respect to biospecimen preparation variables. The value of a biospecimen has at least 4 components. The first component is the intrinsic value of the biospecimen itself for testing. Without the correct, properly collected and preserved biospecimen, accurate and reliable testing cannot be done. The second component is the information value associated with the biospecimen, especially for research. The more information recorded about the life history of the biospecimen and the patient's outcomes in response to particular interventions, the more valuable is the biospecimen. The third component is the rarity value of the biospecimen, which depends on the prevalence of the associated clinical phenotype (eg, rare disease biospecimens). The fourth component is the commercial value of the biospecimen, which depends upon its intrinsic and information values and market demand.

Potential variables were collected from personal patient care and research experience. Definitions and/or notes were supplied for all variables in order to reduce the variability in interpretations and/or to clarify exactly what is the nature of each variable. Eleven negative impact categories were defined.² These negative impact categories describe what may happen if the variable is not collected and attached to the biospecimen.

The manuscript was reviewed before submission by members of the CAP Diagnostic Intelligence and Health Information Technology Committee and other national and international domain experts. These experts are listed at the end of the article.

RESULTS

Four additional clinical chemistry or microbiology-specific required preanalytic variables were identified after reviewing the 170 preanalytic variables in the previous publication² and are listed in Table 1 with their definition/note and examples of potential negative impacts. Table 2 lists the 11 categories of negative impacts that were used in the previous publication.²

COMMENT

Very little field-validated research is available as to exactly what clinical and preanalytic variables should be attached (annotated) to a particular biospecimen. Although we initially tried to identify specific evidence for the recommended variables, there is too little literature support to be useful.²

The complete list of 174 preanalytic variables is not suitable for all institutions. We emphasize that no single universally applicable list can be generated. The choice of the variables will depend on the scope of the collections, while the implementation depends on the capability of the institution's information technology infrastructure to support the collection of chosen variables. Currently, much of the recommended annotation is being captured and entered manually, a process that is often very difficult, error prone, and time-consuming. Obtaining accurate and reliable clinical and research testing results, at the time of collection and in the future, requires biospecimens with high intrinsic and information value. The collection of appropriate and accurate preanalytic variables is essential to providing high intrinsic and information value to biospecimens. We strongly support the extensive use of information technology implementation to automate the collection of preanalytic variables, including other clinical information as

Table 2. Negative Impact Categories With Examples, Reprinted From the Study by Robb et al²

Category	Tissue Example	Blood Example ^a
1. Biospecimen: insufficient for testing or incorrect	A thyroid FNA cell block containing only blood is sent for <i>BRAF</i> testing.	Whole blood previously processed, but only the plasma fraction of blood is now available in the repository, and plasma does not contain adequate leukemic cell DNA.
2. Biospecimen: degraded inventory	A breast tumor that was unfixed for 96 h is sent for hormone receptor testing.	Whole blood received, but after refrigerating the residual whole blood for 5 mo, there is insufficient intact DNA.
3. Biospecimen: source misidentified	A negative lymph node is sent for <i>BRAF</i> testing when a node with metastatic melanoma was the intended substrate. Wrong accession no. used on downstream aliquot (sample) and/or slides.	Whole blood processed, but all the patients with leukemia were in remission rather than at initial diagnosis.
4. Biospecimen: chain of custody and/or history	Information about the biospecimen source and clinical history is mistakenly deleted and no longer available to researchers. Consent comments not saved in database and/or not attached to biospecimen. Complete consent form not attached to biospecimen.	Whole blood received from patients with leukemia that had been processed at an outside facility although it was not clear how long the blood had been stored and at what temperature it was stored before processing at the outside facility: insufficient annotation.
5. History: insufficient or incorrect	A colon cancer sent for microsatellite instability testing is mistakenly described as having been obtained following neoadjuvant chemoradiation, confounding interpretation.	Whole blood previously processed, but the anticoagulant type (eg, EDTA, heparin) had not been recorded.
6. Legal risk	Genetic testing is done on a biospecimen for which the patient did not provide consent.	Whole blood previously processed, but it was not clear whether the subjects had consented to whole-genome sequencing. Another example could be if the receipt at a location is missing, and the biospecimen was actually lost in transit.
7. Patient: incorrect identification	Biopsy of tumor mistakenly labeled as belonging to a different patient.	Whole blood previously processed and one of the samples had the diagnosis “PCD” in accompanying paperwork. Because PCD was an ambiguous term (eg, plasma cell dyscrasia, primary ciliary dyskinesia), this sample did not meet the needs of the investigator.
8. Quality: testing result issue	False-positive <i>HER2</i> test results are recorded. Histologic slide is mislabeled. Incorrect weight is attached to biospecimen.	Whole blood previously processed with karyotyping done at site C, which was not a CLIA-certified laboratory. Because the annotated karyotype results from site C were not clinical-grade results, the samples did not meet the needs of the investigator.
9. Quality: missing quality indicator	Total time in fixative is not available.	Whole blood previously processed, but neither blood smears nor blast counts were checked before processing to determine whether the samples met study inclusion criteria regarding minimum proportion of leukemic cells.
10. Resources wasted	An incorrect panel of biomarker tests is done on a small biopsy sample, and the tissue is depleted as a result. Cancer aliquot (sample) is too large for shipping and has to be trimmed.	The samples had not been annotated with information on how the blood was processed.
11. Operational: investigator confidence in biobank	Billing for biospecimens not properly explained.	While a database of available biospecimens can be made accessible, a contact person is needed to clarify data elements and to promote efficient use of tangible and information resources.

Abbreviations: *BRAF*, proto-oncogene B-Raf gene; CLIA, Clinical Laboratory Improvement Amendments; EDTA, ethylenediaminetetraacetic acid; FNA, fine-needle aspiration; *HER2*, human epidermal growth factor receptor 2.

^a100 Patients for a Study of Leukemia Cell DNA.

needed, and subsequent annotation of the biospecimens with those variables, as well as biospecimen processing variables (eg, quality indicators). In most instances, manual collection with subsequent annotation is simply not an alternative.

It is clear that the risk of potential adverse events will increase if variables are not attached to the biospecimen. The cumulative adverse impact will depend on which variables are omitted. The 11 negative impact categories

with specific examples are useful for understanding the potential risks of suboptimal biospecimen annotation.²

The development of functional technical specifications for how these enumerated variables might be implemented into an existing software system is a critical next step. The implementation into an existing software system is necessary in order to achieve measurable improvements in biospecimen quality and patient care. The CAP Diagnostic Intelligence and Health Information Technology project

team will begin practical implementation discussions with the international scientific community, information technology vendors, and other stakeholders. These discussions are intended to disseminate the functional specifications of preanalytic variables and to determine what type of input is needed from a wide range of interested parties. Terms and definitions must be internationally aligned for the practical exchange of biorepository inventory information. Measurement units for preanalytic variables, where appropriate, must be part of that effort.

Feedback is welcomed on the 174 enumerated variables, as well as the prioritization of variables in either clinical or research settings. This effort represents the beginning of a longer, more involved process.

There is tremendous benefit from explicitly proposing a set of standard data elements relevant to biospecimen quality in an era of precision personalized medicine. While much of the current effort is based on expert opinions, our expectation is that evolution of this effort will crystallize those variables that have the most impact on health care outcomes and research findings. We anticipate that relevant stakeholders will join and engage in an effort to collate hard data on the usefulness of the proposed elements to concretely identify the most relevant and practically

implementable variables. Finally, significantly more biospecimen research is needed to develop supportive evidence for the relative importance of the 174 preanalytic variables already identified.

We thank the following domain experts for their valuable review of the manuscript: Fay Betsou, PhD, HDR (experience with the International Society for Biological and Environmental Repositories and the Integrated BioBank of Luxembourg); Victoria Blanc, PhD (experience with the College of American Pathologists Biorepository Accreditation Committee, Office of Research, University of Michigan Medical School, Ann Arbor); Latarsha Carithers, PhD (experience with the National Cancer Institute's Biorepositories and Biospecimen Research Branch, Rockville, Maryland); and Sarah Dry, MD (experience with the College of American Pathologists Biorepository Accreditation Committee, Department of Pathology and Laboratory Medicine, University of California, Los Angeles). These experts have experience with activities similar to those described herein. Their input was based on personal experience and does not represent the view of the organizations with which they are affiliated.

References

1. Dash RC, Robb JA, Booker DL, et al. Biospecimens and biorepositories for the community pathologist. *Arch Pathol Lab Med.* 2012;136(6):668–678.
2. Robb JA, Gulley ML, Fitzgibbons PL, et al. A call to standardize preanalytic data elements for biospecimens. *Arch Pathol Lab Med.* 2014;138(4):526–537.