Developing a data repository of standard concussion assessment clinical data for research involving college athletes

Arthur C. Maerlender  
*University of Nebraska-Lincoln, amaerlender2@unl.edu*

Jennifer Mize Nelson  
*University of Nebraska–Lincoln, jnelson18@unl.edu*

Julie A. Honaker  
*University of Nebraska-Lincoln, julie.honaker@unl.edu*

Follow this and additional works at: [http://digitalcommons.unl.edu/cbbbpapers](http://digitalcommons.unl.edu/cbbbpapers)

Part of the *Behavior and Behavior Mechanisms Commons, Nervous System Commons, Other Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons, Other Neuroscience and Neurobiology Commons, Other Psychiatry and Psychology Commons, Rehabilitation and Therapy Commons*, and the *Sports Sciences Commons*

[http://digitalcommons.unl.edu/cbbbpapers/1](http://digitalcommons.unl.edu/cbbbpapers/1)

This Article is brought to you for free and open access by the Brain, Biology and Behavior, Center for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Center for Brain, Biology and Behavior: Papers & Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Developing a data repository of standard concussion assessment clinical data for research involving college athletes

Arthur Maerlender,1 Jennifer Mize Nelson,1 and Julie A. Honaker2

1 Center for Brain, Biology, and Behavior, University of Nebraska–Lincoln
2 Department of Special Education and Communication Disorders, University of Nebraska–Lincoln

Corresponding author — A. Maerlender amaerlender2@unl.edu C86 East Stadium, University of Nebraska–Lincoln, Lincoln NE, 68588-0156

Abstract
In sports concussion research, obtaining quality data from a sufficient number of participants to reach statistical power has been a particular problem. In addition, the necessary requirements of accessibility, informed consent, and confidentiality must be met. There is need to develop more efficient and controlled methods for collecting data to answer research questions in this realm, but the ability to collect and store these data in an efficient manner at the local level is limited. By virtue of their training, neuropsychologists can play a key role in improving data collection quality. The purpose of this paper is to describe a data repository that has been developed in the context of a university sports medicine concussion management program that includes baseline and postinjury data from student athletes. Diagnostic information, basic health information, current symptoms, neuropsychological test data, balance and vestibular data, and visual processing data are currently included in the standard of care for athletes; however, the process described need not be limited to these types of data. While a national traumatic brain injury (TBI) data repository has been developed by the National Institute of Health (NIH), local repositories have not yet become common. Thus, the description of this project is of value at the local level in the United States and internationally.

Keywords: sports concussion, data repository, clinical standard of care, data integrity, data collection standards

In sports concussion research, obtaining quality data from a sufficient number of participants to reach statistical power has been a particular problem. In addition, the necessary requirements of accessibility, informed consent, and confidentiality must be met. Indeed, there is a need to develop more efficient and controlled methods for collecting data to answer research questions in this realm (Maas et al., 2011; Tossetti et al., 2013). While standards for biological (i.e., biospecimens such as blood, plasma, or tissue) repositories are established (see University of British Columbia, BC Cancer Agency, Research Ethics Board, UBC BCCA REB, 2011), the need for local data collection and storage standards for nonbiological data (i.e., clinical assessment data such as current symptoms, neuropsychological and balance test data, and health status case history) are not in existence for local entities. Indeed there is a lack of local sport concussion data repositories utilizing clinical assessment data, and, to our knowledge, this is the first reported local data repository specifically designed for sports concussion research.

Utilizing a standardized approach to collect concussion assessment data is necessary not only to make informed clinical decisions, but to ensure valid and reliable data from a controlled environment for optimal research analysis. For instance, we recently compared several years of baseline neuropsychological data collected in large group testing sessions to data collected under more controlled conditions. The reliability increased significantly under the more controlled conditions from .59 to .74 (Cronbach’s alpha; Fisher’s z-test p = .005: Higgins, Maerlender, & Caze, 2016).

Following standardized and well-controlled procedures is particularly important for psychometric data whose inferential capacity is tied directly to appropriate administration. This includes neuropsychological data and also holds true for criterion-referenced data such as symptom checklists, observational balance data, physiological data (e.g.,
videonystagmography), and diagnostic information from other allied health professionals. Therefore, neuropsychologists working closely with other allied health professionals on a team are in the best position to lead the way in these endeavors by virtue of training and experience. Standardized data collection includes administration by competent individuals according to published procedures (see American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, AERA, APA, & NCME, 2014). Such standardization will provide an opportunity to utilize the clinical data for research purposes, through secondary research analysis. Further, standardized data sets may provide answers regarding program quality such as (a) does the data provide an answer to the clinical question; (b) are all elements of the standardized protocol necessary, and/or are other assessments needed?

The term “secondary data” refers to data originally collected for a purpose other than the current research purpose. Within a data repository, a researcher may seek to use materials left over from a clinical procedure, or materials collected from an earlier research project. Datasets can be compared by various groupings to determine systematic group differences (i.e., sport teams, gender, years, etc.). Researchers who are conducting other specific studies may then have access to the clinical dataset as secondary data to help inform their study (e.g., understanding gender or team differences on assessment measures to facilitate the design of a prospective study). Perhaps most importantly, providing a set of data points may allow researchers conducting other studies to minimize their assessment and data requirements.

By obtaining consent for secondary uses at the time of baseline concussion testing, the subjects (e.g., in the case of student athletes) do not need to be re-consented for every possible use of the data, thus minimizing their burden. As examples, the BIG10-Ivy League Traumatic Brain Injury (TBI) Research collaboration conducts an annual surveillance project to identify concussion rates. A data repository of demographic information allows the surveillance project to reduce the additional data collection requirements. Similarly, involvement in a research project that overlaps with standard of care testing would only necessitate requesting informed consent for those measures that are not already included in the data repository of standard of care measures. Thus, the researchers could request secondary data analysis rather than asking the student athletes to complete redundant measures.

The purpose of this paper is to describe a data repository that has been developed in the context of a United States university sports medicine concussion management program that includes baseline and postinjury data from student athletes. Within this U.S. university sports medicine concussion management program, diagnostic information, basic health information, current symptoms, neuropsychological test data, balance and vestibular data, and visual processing data are currently included in the standard of care for athletes. The repository process described here need not be limited to these types of data. An important aspect is that this information should be part of standard clinical procedures and not experimental procedures that would require separate consents.

In the United States, concussion management is now currently mandated in all 50 states, so this data repository may have applicability to a wider range of venues within the United States. Data repositories for more general access have been developed by the National Institute of Health (NIH) in the case of the Federal Interagency Traumatic Brain Injury Research (FITBIR) system, but local repositories have not yet become common.

The International Compilation of Human Research Standards (United States Department of Health and Human Services, USDHHS, 2016) enumerates over 1,000 laws, regulations, and guidelines that govern human subjects research in 120 countries, as well as the standards from a number of international and regional organizations, including privacy and data collection issues. This document is based on the World Medical Association’s set of Ethical Considerations regarding Health Databases (2002, in World Medical Association, WMA, 2016) that is still regarded as a standard across the globe. That declaration defines de-identified data and its appropriate use, data integrity, and documentation. The principles and guidelines in this document are consistent with the broader scope of data collection worldwide. However, it is important to note that local entities (even within the United States) create their own specific rules and procedures that differ from institution to institution, both within and between countries.
Process for setting up a data repository

Several components and processes may need to be delineated to a research governing committee (e.g., U.S. institutional review board, IRB) for approval and use of human subject research. The IRB is specific to the United States and is a research ethics committee commonly used in medical research to review and approve human subject research, to protect the welfare of human research participants. The necessity IRB review is institution specific regarding research compliance regulations for granting approval to allow human subject’s research and ultimately the data to be included in a data repository. The authors’ experience is that data of this nature require IRB approval, which may be specific to our institution and U.S. research compliance regulations. Some of the elements included in the review process are specific to our location, but some are generally applicable. These components include (a) the purpose of the repository; (b) governance and accountability; (c) informed consent procedures; (d) standard operating procedures; (e) sustainability and succession planning; and (f) intellectual property (UBC BCCA REB, 2011). It is imperative to clearly articulate the nature of the data repository and to work closely with the local athletic department and IRB administrators to facilitate the initial approval process for the local data repository. Potential issues that researchers might encounter when initially setting up the repository are the need for parental consent for minors and the process for contact information requests to obtain parental contact information. An additional issue is the need for a process to re-consent or provide a reminder of consent for students who sustain a concussion regardless of whether they consented/assented to the original data repository. This is an important discussion with IRB administration given the potential time from the student athlete’s initial consent/assent.

Governance and accountability of the repository

Management of the repository must be specified, including relationships between data collectors and data managers. In this scheme, the manager of the repository is seen as an “honest broker” who has no vested interest in the research or project that generates the data. This has important implications for managing identifiers and the potential for linking identifiers. The manager represents the institution and thus has ultimate responsibility for proper storage and use. Both standardized processes and decision-making procedures should be specified. Typically an advisory board of disinterested scientists should review requests for access and withdrawal. These steps are necessary to inform the decision-making process for deposit, withdrawal, and access to the data repository.

Informed consent for data inclusion in the repository

During the consent process, a clear distinction should be made between consent for research participation and consent for any clinical procedure or test. The consent should then specify the permission to use the project-specific subject’s data for future research in which their identifying information will be removed (de-identified). Participants may also be asked permission to be contacted for later studies.

Standard operating procedures (SOPs) for the repository

There should be a procedures manual that includes the data elements, procedures, and documentation requirements and provides continuity of function and process over time. Table 1 lists the recommended elements for SOPs.

Sustainability and succession planning for the repository

Obtaining institutional management awareness and support is necessary, while a funding or financial plan is needed to assure sustainability into the future. If there is a potential limit to the data use and collection, the “fate” of the data should be described.

**Table 1.** Recommended elements of standard operating procedures (SOP).
(I) Scripts and procedures for obtaining consent and ensuring that proper consent has been obtained;
(II) Data/specimen acquisition policies;
(III) Fidelity procedures and ensuring specimen/data integrity, including data entry and checking;
(IV) Subject coding and de-identification procedures;
(V) If biological samples are included, identify rules for transporting specimens/data;
(VI) Ensuring proper data storage environments;
(VII) Maintaining confidentiality;
(VIII) Screening potential researchers in terms of suitability/eligibility;
(IX) Overall quality control processes.

**Intellectual property of the repository**
There should be a clear set of rules around handling of intellectual property and how relationships with funding sponsors will be handled. If funding for the resource is coming from some external sources, the organization and IRB must understand the conditions, if any, that accompany such commitments and whether there are written agreements in place. Legal counsel may be needed.

**University of Nebraska–Lincoln (UNL) athlete data repository: A case example**

**Purpose**
The purpose of the UNL Athlete Data Repository is to safely and securely store standard of care (SoC) and research data collected with UNL varsity student athletes, to allow for secondary data analysis, and support longitudinal studies.

Data collected through SoC are clinical data and thus identified data. To facilitate anonymity, the repository establishes a coding and de-identification system to securely store the data and enable researchers to access only de-identified data with appropriate IRB approval for research purposes. Other researchers collecting varsity student athlete data are also invited to store their data in the repository; consents for new research studies may include a checkbox to signify agreement of storing the data in the repository at the conclusion of the study. Thus, the repository includes both SoC and research measures for a comprehensive repository of data on varsity student athletes.

A particularly valuable use of this process is for assessing local procedures. By examining the outcome of repeating baseline neuropsychological testing every two years, we have shown that this is not a necessary practice, and baseline testing once at college entry is sufficient (Maerlender & Molfese, 2015: “Repeat Baseline Testing in Collegiate Athletics”). A second paper in development demonstrates the increased reliability obtained by enforcing strict procedural and administration guidelines when administering baseline tests. Another project involves the incremental clinical utility of combining multiple vestibular, visual, and neuropsychological tests in both baseline and postinjury assessments. The university also participates in a multisite collaboration for surveillance of concussion incidence. By obtaining this type of consent, more extensive data can be included in the simple incidence data to enrich understanding within and across institutions about the rates and proximal causes of injury. Finally, as the university adds clinical tools, data from electrophysiology, functional magnetic resonance imaging (fMRI), and biosalivary assays will enhance clinical analysis. The ability to easily combine all forms of data with basic demographic information and sport-specific information is a key component of the data repository.

Throughout the development of this project, an ongoing tension between data platform and data structure was discussed. On the one hand, data were accumulating in various formats within each lab, primarily in Excel spreadsheets. We felt it was important to let the data manager define the structure of the data before determining the platform. Other data repositories on campus utilized programmers working in language such as SQL. Although powerful, this made data access and entry dependent on programmers. The complexity of the programs made sustainability and end-user capability challenging. Using a user-friendly “off-the-shelf” product allows for ease of use and flexibility for making changes.

**Operating procedures**

**Data/specimen acquisition**
Data considered part of the SoC for Athletic Medicine services are collected at two primary time points (baseline and postinjury assessments). The baseline data (pre participation) are collected during the athletes’ first season and thereafter as frequently as indicated by recent National Collegiate Athletic Association (NCAA) guidelines (2015). Athletes are first approached about the repository during the baseline time point and consented at this time point. Postinjury assessments (as recommended by the UNL SoC protocol; e.g., re-assessment when symptom free)
are also captured. Athletes who present at post injury are checked for having consent on file. Those who have not been consented are invited to consent; the consent document includes language to access all previous SoC data as well. Upon consenting, the previous data in addition to any data collected on the day of consenting are added to the data repository. An example of informed consent is presented in the Appendix.

As noted above, the majority of SoC data are collected in the athlete’s freshman year during baseline assessments (as above). For underage athletes, both student athlete assent and parental consent are required (in Nebraska, athletes under the age of 19 years require parental/legal guardian consent and athlete assent). Procedures for streamlining this process have been developed cooperatively with the Department of Athletics.

In addition to varsity student athletes, the repository may also include nonvarsity athletes (i.e., club sport athletes). Within the UNL repository, all nonvarsity athletes (i.e., club sport athletes) who also provide baseline and postinjury SoC data are also included following similar procedures. A unique general universal identification (GUID) specific to nonvarsity athletes are assigned as each athlete is consented. This ID assignment and the actual depositing of the data are completed by the data manager.

In order to provide high-quality data for clinical and research purposes, it is critical that clinical data are acquired under the supervision of the appropriate clinical specialist. Verification of competence of administration and fidelity of procedures is necessary for empirical use. Within our facility, all SoC is conducted by appropriately trained and supervised graduate students who have education in assessment practices and specific test administration. Supervision is provided by licensed (and board certified) clinical practitioners to insure reliable and valid results. Data are interpreted by the clinical team working with Athletic Medicine. The clinical team then provides a dataset with the SoC data to the data manager (honest broker), who then codes and de-identifies the data to be stored within the repository. The unique code provided by the data manager is the primary source of identification for all records in the repository.

Data access and extraction

The data stored within the repository is strictly monitored, and only those investigators with appropriate IRB approval are able to access the specific data that was specified in the proposal. As a first step to acquiring the data, all investigators interested in utilizing the data for secondary data analysis must first submit a brief proposal specifying the type of data requested and the aims of their project. A team of faculty members then review the proposals and provide a letter of support or refusal to the IRB to use the data. At this point the IRB may proceed with an expedited review and waiver of consent given the secondary data analysis and previous consenting process for data storage within the repository. With IRB approval, the data manager then provides de-identified data to the researcher, while maintaining a record of the request, proposal, IRB and description of the data provided. Appropriate attribution is required in all publications and scientific documents deriving from the data repository.

Because the data deposited in the repository change over time, succeeding consents are stored together with a list of athletes and the consent signed. In this way, it is possible to know exactly what data the athlete has consented to for storage in the repository. This list is held by the data manager and is updated by staff.

Summary

This paper describes the general goals, components, and procedures for establishing and maintaining a data repository of athlete standard of care data in a university setting. It then describes one approach developed to create a data repository that includes sports concussion management data. In this case, the repository is to systematically store both routine clinical/standard of care data, as well as individual research study data, from student athletes who undergo routine baseline and postconcussion assessments. The descriptions here are broadly applicable for other uses. A major goal of this initiative was to collect these data in a manner that is as unobtrusive to the student athlete as possible. The intention was that the repository will allow for unique and innovative studies to improve the health and well-being of the student athletes.

As noted, attention to data collection is a key component of the process outlined. A unique challenge is the intersection of clinical and research goals inherent in this approach. The quality of data collection is often assumed without verification in data collection schemes. Indeed, clinical reality often obviates one’s ability to administer instruments in a strict,
standardized manner. The intention is to identify data at acquisition that is not up to research standards so that it is not included in empirical studies. The twin goals outlined (both research fidelity and clinical utility) are familiar grounds for clinical neuropsychologists whose extensive training and interests often find them involved in data acquisition and analysis. This does not mean that individuals in other disciplines are not so involved, but this approach provides neuropsychologists a unique opportunity to exercise their skill-set.

The motivation for this project stems from the conclusion that clinical sports concussion management, and the necessary research associated with it, are at a point in time where procedures need to be improved in order to obtain clinical and research data that can drive best practices and begin to answer important questions about traumatic brain injury (TBI). With the explosion of off-the-shelf tools, the lack of rigor in administration (and interpretation) continues to limit the value of research (for instance, see Meier et al., 2015). This data repository seeks to establish a process for obtaining data of sufficient quality to be used for research purposes, including program evaluation. Due to the variability inherent in local human subjects governing bodies, specific details may not be applicable. The intent is to provide a broad framework to initiate those discussions.

Acknowledgments — The authors would like to thank the staff of the University of Nebraska–Lincoln (UNL) Institutional Review Board (IRB) and the Department of Athletics for their encouragement and support of this repository.

Disclosure statement — No potential conflict of interest was reported by the authors.

ORCID — Arthur Maerlender ORCID: 0000-0001-9740-6354

References


Appendix

Informed consent form (IRB#)

(1) Study Title: Athletic Data Repository
(2) Purpose of the Study: As part of your medical standard of care for concussion management, you participate in several baseline assessment procedures within the <name>. If concussed, you complete these same assessments to determine your recovery and readiness to return to play. The purpose of this project is to allow your data from these standard of care procedures to be used in research studies in a manner that does not identify you. Your consent or assent (together with parental consent) will allow <name> to store and use your data for research without identification and without further consent.
(3) Inclusion Criteria: You are currently a <name> student-athlete undergoing standard of care assessments for concussion management.
(4) Study Procedures: We are seeking your permission to save the information recorded during your baseline and postinjury concussion assessments in the Athletic
Data Repository. We have outlined information on the types of assessments you may have participated in below. Please place your initials on the line indicated if you give us permission to confidentially save that portion of your information in the Athletic Data Repository.

a. **Background information**: Information about your demographics (sex, race, age); dates of assessment; academic performance; neurological, developmental, and psychiatric health given known relevance for incidence and outcome of concussions; number of previous concussions or head injuries; time between concussions; concussion symptom duration; and current medications are collected.

b. **Neuropsychological data**: Data from neuropsychological tests taken at baseline and post-injury will be included (i.e., Immediate Post-Concussion Assessment and Cognitive Testing: ImPACT; C3 Logix tests, paper and pencil tests as indicated for your care).

c. **Balance and vestibular clinical assessment measures**: To assess balance and vestibular functions, including vision, several tests are conducted by trainers and by the Balance Disorders Lab. Your eye movements are tracked with special goggles to monitor your inner ear balance function, in addition to parts of your brain that monitor your ability to move your eyes and follow a target. A sensor on your head records your head movements while viewing a computer screen. A forceplate measures your body’s sway during challenging standing balance activities; and a special i-tablet may be used to measure sway as well. Additionally, responses to balance and head injury history questions provide additional information.

d. **Concussion information**: Should you be concussed during sanctioned Athletic Department activities, information about your injury will be included. Items will be entered such as the sport you were playing, whether or not it was a game or a practice, the mechanism that caused the injury, the situation and how the injury occurred, the length of time you were out of activities, whether or not a foul or penalty was called. We will also like to include scores from the Standardized Concussion Assessment (SCAT-3) and any other medical data relating to your injury. This information will help us to better understand the nature of concussions, their causes and consequences.

(5) **Risks**: There are minimal risks associated with consenting for your data to be deposited in the Athletic Data Repository. The primary risk, albeit unlikely, is a risk to privacy and confidentiality of stored information should there be a breach of the data repository.

(6) **Benefits**: There are no immediate direct benefits to participants as a result of storing their data in the data repository. However, the data repository will allow standard, clinical data to be used for research purposes aiming to better understand student-athlete health and well-being, including concussion incidence, trends, management, and length of time to recovery. To the extent these findings influence team and medical care decisions, participation may have benefits of improving the safety of your sport.

(7) **Compensation**: You will receive nothing of monetary value for your involvement.

(8) **Privacy**: Your clinical standard of care data will be made available to your medical team in an identifiable form. Once your data are provided to the medical team it falls under the Athletic Department confidentiality practices.

All data you give us permission to include in the Athletic Data Repository will be treated confidentially. All information will be kept in locked drawers or stored electronically on a secure University server with password protection and firewalls. Only this consent form contains your name. Identifiable information (such as the consent form) will be stored separately from all data. Each participant is assigned an alphanumeric ID, and data stored within the repository will be coded only by the alphanumeric IDs. The list linking the alphanumeric IDs with names is stored separately from any data, only accessible to select members of the research staff associated with the Athletic Data Repository. This list will be kept for the life of the repository.

Data released from the repository for research purposes will be de-identified, not including any meaningful associated IDs.

If you choose to include your data in the repository, your data may be used in articles published in scientific journals or presented at scientific meetings. Neither your name nor any photo or video image of you will be used within such publications or presentations.

Additional research studies may arise within <name>. If you are interested in learning about these future studies, you can choose to give our research team permission to save your contact information in a secure file and to try to contact you in the future.

Initiaing indicates your consent to allow us to save your contact information (e.g., name, email, phone number) in a secure file and to contact you in the future for additional research studies.

**Parent or Legally Authorized Representative**

Initials: __________

Participant Initials: __________

If you decide to withdraw from this data repository project after agreeing to participate, we will delete your contact information and will refrain from
contacting you about future research opportunities within <name>. However, your de-identified data collected prior to your withdrawal will remain in the repository. Withdrawal from the project can be accomplished by contacting the investigators, or if they contact you in the future regarding another research study within <name>, you can inform them that you would like to withdraw and have your contact information deleted.

(9) Questions: You may ask the investigators questions and have those questions answered before agreeing to participate, or at any time during the life of the data repository. If you have questions, you may call the investigators at the phone numbers listed below. If you have questions about your rights as a research participant that have not been answered by the investigators, or if you wish to report concerns about the data repository project, please contact the University Institutional Review Board at (xxx) yyy-zzzz.

(10) Signature: You are free to decide not to participate in the data repository or to withdraw at any time without adversely affecting your relationship with the investigators or University. A decision not to participate, or withdrawal from the repository project, will not affect your relationship with University Athletics, nor will it affect return-to-play decisions. A decision not to participate will not result in any loss of benefits to which you would otherwise be entitled. Your signature below (and that of your parent if you are under 19 years of age) indicates that you have voluntarily decided to participate, and have read and understand the information presented. You may have a copy of this consent form to keep.

Consent/Assent to participate in the <name>Athletic Data Repository:

Participant Printed Name ________________________________  ______/______/_________

Participant Signature ________________________________  Date Signed  ______/______/_________

Researcher Signature ________________________________  Date Signed  ______/______/_________

Parent or Legally Authorized Representative Permission
(only for athletes under the age of 19)

Legal Name ________________________________  ______/______/_________

Signature ________________________________  Date Signed  ______/______/_________

Relationship to participant: __________________________________________________________

Names and phone numbers of investigators: