Boston Medical Center Clinical Data Warehouse
Procedure Manual

PURPOSE

Created in 2005, the Boston Medical Center (BMC) Clinical Data Warehouse (CDW) houses clinical information from BMC's electronic health records (EHR) and other health system-related data streams. For projects with IRB approval, CDW analysts can provide extracts of data from the CDW for research teams to answer research questions or identify cohorts for research studies. The CDW can also provide count data for feasibility analyses or proposal/grant preparation.

The purpose of this Procedure Manual is to communicate the policies and procedures by which the CDW fulfills its mission to provide investigators with high quality, reliable data and to provide guidance for investigators requesting CDW data for research. The manual, which will be updated regularly to communicate evolving policies, outlines policies and procedures in three categories:

Requesting Data & Accessing CDW Services.

Policies regarding who can request CDW data, for what purposes, and with what approvals. This section includes a description of CDW services, an outline of the request process, and expectations of research teams.

Billing & Invoicing.

Guidelines concerning fees for CDW services and CDW billing practices.

CDW Team Operations & Management.

Standard operating procedures to maintain quality and confidentiality of data.

v1.0
Last Updated: March 2021

Table of Contents

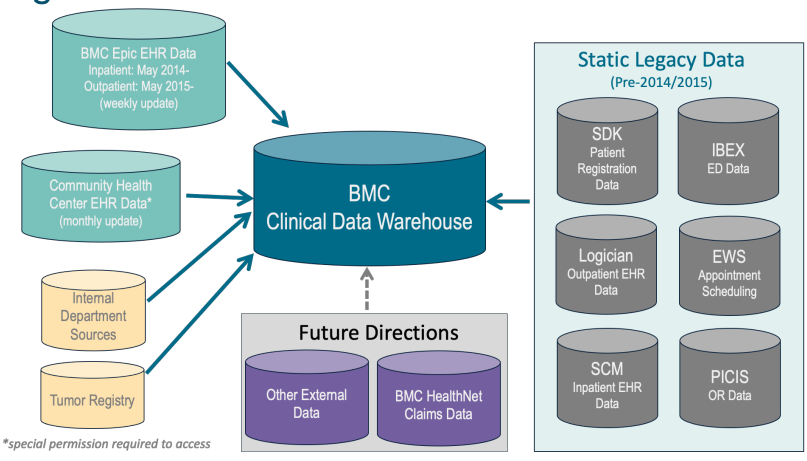
ABOUT THE CLINICAL DATA WAREHOUSE

The BMC Health System is a healthcare network that includes BMC, affiliated community health centers (CHC), and BMC HealthNet — a non-profit managed care organization that provides health insurance coverage primarily through Medicaid contracts. The CDW consolidates patient data from the Health System's EHR and other health system-related data streams to create a repository of data to support research. The CDW is run by a team of analysts that report to BMC's Associate Chief Medical Officer for Research and is overseen by BMC's Chief Medical Office. An Advisory Committee assists CDW leadership in determining the scope of CDW services, assessing whether the CDW is fulfilling its mission to provide high quality data efficiently for research, and identifying opportunities for improvement and growth.

Data are collected and consolidated into the CDW on a regular schedule, but not in real time. Data from many of the systems that contribute to the CDW can be cross-referenced or linked using unique patient identifiers. BMC EHR data are pulled from Epic. CHC data are pulled from OCHIN (a different version of Epic used by a national network of community health centers). Available clinical data include patient demographics, data collected during health care encounters (in-patient admissions, offices visits, emergency department visits, surgical procedures, telemedicine, etc.), diagnoses, procedures, documentation flowsheets, lab values, medications, and clinical notes. Most data that are recorded in Epic are captured in the CDW. In some instances, however, certain data elements that are visible to frontline Epic users may be generated at the point of care and may not be captured within the CDW.

The CDW can provide count data for feasibility analysis or proposal/grant preparation without IRB approval (so called prep-to-research). With IRB approval, analysts can provide extracts of individual-level data from the CDW for retrospective or prospective research studies. For some data streams, investigators may need specific permission to access those data from the CDW, including OCHIN data from CHCs.

Figure 1. Data Housed within the CDW



In addition to data from Epic and OCHIN, the CDW includes data from various data systems that are no longer in active clinical use (legacy systems). These legacy systems include PICIS (operating room data), IBEX (emergency department data), SDK (patient registration data), SCM (inpatient EHR data), Logician (outpatient EHR data), and EWS/IDX (patient scheduling data). Legacy systems are not actively maintained; data from them can be cumbersome to extract and are not monitored for quality.

REQUESTING DATA & ACCESSING CDW SERVICES

Scope of CDW Services

With appropriate approvals, researchers may request data through a CDW analyst but cannot access CDW data on their own. The CDW team provides the following services:

Help Formulating a Data Request

A well-defined data request is critical to obtain accurate, reliable and useful data. CDW analysts are experienced in helping investigators refine data requests, determine what types of data are available and appropriate to address the research question(s), and articulate the request in the appropriate format within the CDW Data Request Form. The CDW analysts will consult with research teams interested in requesting data from the CDW. **An initial, one-hour consultation is provided to all studies free of charge.**

Estimate the Cost of a CDW Request

The CDW analysts are available to make cost estimates for their work. Investigators must include the cost of CDW services in their grant proposal budgets. Payment for CDW services can be based on an hourly rate or on an estimated a percent time for analyst work. Investigators interested in paying for services via an FTE structure need to pre-arrange this with the CDW prior to the project's start. Estimates are subject to change if the project deviates in scope after the initial estimate, or if there are unforeseen complications in obtaining relevant data fields. CDW analysts will communicate to the investigator when there is a significant change in the cost estimate (±20% deviation) to avoid unexpected charges.

Provide Count Data for Feasibility Analysis, Prep-To-Research, or Proposal/Grant Preparation

The CDW team is available to provide investigators with count information (i.e. aggregated data, not supplied at the individual patient level) to prepare for a research study. The CDW provides three free hours of count request services to investigators with the expectation that funding for further CDW services are included in proposal budgets. Requests of count data (e.g. the number of hospitalized patients with pneumonia in 2018) do not require IRB approval. Counts fewer than six cannot be reported due to privacy concerns.

Provide Extracts of Data Housed in the CDW

With appropriate approvals, CDW analysts can extract data sets with individual-level data for investigators. IRB approval is required before the CDW can begin working on the following types of data extract requests:

> Cohort identification: The CDW can assist research teams in identifying cohorts of patients that meet study eligibility criteria. For example, the CDW can provide a report of patients

that meet study inclusion and exclusion criteria to the research team to facilitate subject enrollment.

Data sets: The CDW provides de-identified data sets, limited data sets, and identifiable data sets, either retrospectively or prospectively. For example, a research team studying acute health care utilization among patients with neck abscesses may request a data set that contains patient demographics; emergency department visit dates; in-patient admission and discharge dates; operating room information; and lab values such as the value and date of the most recent white blood cell count.

Linked data sets from multiple data sources: The CDW provides data sets that combine data from multiple systems. For example, the CDW can crosswalk BMC EPIC clinical data with CHC OCHIN data for patients with health care encounters at BMC and CHCs.

Out-of-Scope Services & Other BUMC Resources to Support Researchers

At present, the CDW does not assist with the following tasks:

Study design and IRB application

Although CDW analysts are available to help refine data requests and operationalize research questions, analysts do not provide methodologic consultation to help investigators develop their research questions, study designs, or statistical analysis plans. Analysts do not provide assistance developing IRB applications beyond a discussion of which CDW data elements to include. For assistance with study design and IRB preparation, investigators can access the following resources:

- BU Clinical & Translational Science Institute (CTSI): https://www.bu.edu/ctsi/training-education/seminars-and-workshops/open-study-design-and-statistical-analysis-consultation/
- BUMC Clinical Research Resources Office (CRRO): https://www.bumc.bu.edu/irb/contact-us/

Data management, programming, and analysis

While the CDW works with investigators to provide data in a user-friendly format, it does not offer data cleaning, data management, or data analysis services. Generally, the CDW does not conduct any calculations nor derive variables beyond simple counts. The CDW does not calculate descriptive statistics (mean, minimum, maximum, etc.) or do statistical programming. Data are provided in a raw extract format. For assistance with data cleaning, management, or analysis, the following contacts may be helpful:

- BU CTSI's Biostatistics, Epidemiology & Research Design (BERD): http://www.bu.edu/ctsi/support-for-research/biostatistics-and-data-management/
- Biostatistics and Epidemiology Data Analytics Center (BEDAC): https://www.bumc.bu.edu/busm/research/cores/bedac/

Non-research related projects

The CDW is intended for research purposes only. For all other data request purposes, please submit a ServiceNow ticket at bmc.service-now.com.

Expectations of Investigators and Research Teams

The graphic below depicts expectations of investigators and research teams throughout the CDW data request lifecycle.

Prior to CDW Data Request Submission	During Request Submission Process	After Request Submission	Upon Receipt of CDW Data
• Know your research question. • Have a general understanding of the data needed to answer your research question. • Begin acquiring necessary approvals (IRB, CHC, etc.). • Make a plan to pay for CDW services by including costs in grant budgets or speaking with your department leadership.	• Disclose research question and intent. • Provide detailed definitions and ICD/CPT codes. • Ensure data requested from the CDW are reflected in your IRB protocol. • Request all data elements needed to create your derived variables (i.e. variables of analytic interest created by calculation or categorization). • Provide billing account information. • Discuss your request with a CDW analyst during a free, one-hour consultation.	• Expect some back-and-forth with the CDW as analysts hone your request into a data extract. • Let the CDW know when IRB approval is granted. The CDW does not begin a project until notified of IRB approval *by the research team*. • Understand where CDW services end and analysis begins. • Expect your data set to be provided approximately 4-6 weeks after request submission or notifying the CDW of IRB approval.	• Safeguard the data according to applicable policies. • Review data and reach out to the CDW with questions or concerns within two weeks. • Submit a new CDW Data Request Form if you need added components (new variables; new timeframes; etc.). • Submit a new CDW Data Request Form 30 days prior to when data extract is needed (if you need a subsequent or recurring file of the exact data set).

Process to Request Data from the CDW

The CDW has processes for investigators to access data. The following is an outline of these processes with recommendations on how to obtain data most efficiently at the lowest cost.

Step 1. Consult with a CDW Analyst

An initial, one-hour consultation is provided to all studies free of charge. All investigators, particularly new investigators or those with early experience working with EHR data, are encouraged to set up a consultation with a CDW analyst to discuss feasibility, scope, cost estimates, and how to optimize the data request. This consultation can take place prior to your study's IRB approval – and can be helpful to determine what data should be included in your IRB submission. Email cdw@bmc.org to request a consultation.

Step 2. Obtain IRB Approval for Your Study

All CDW requests other than simple counts must have IRB approval before the CDW can begin working on a request.

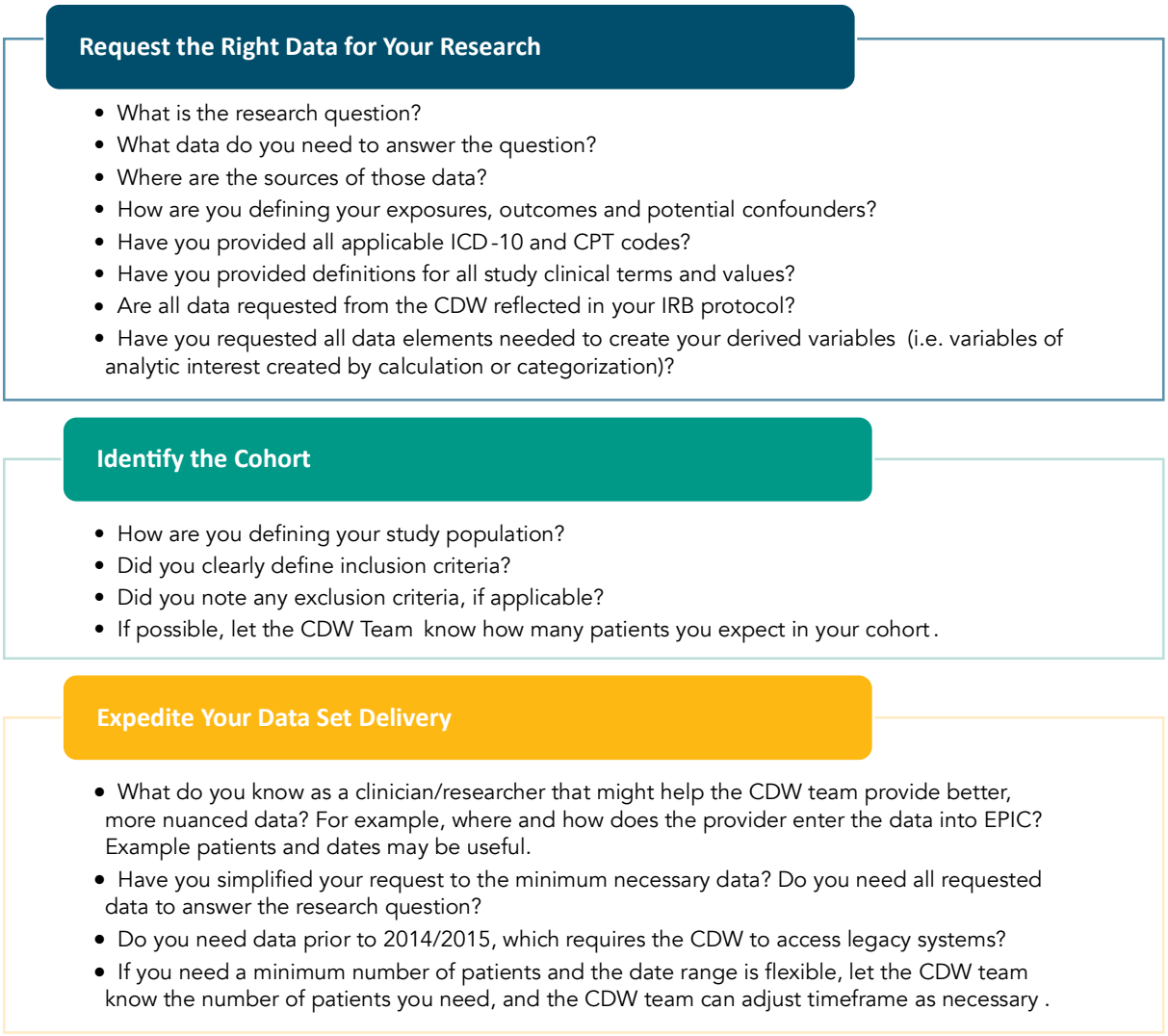
- A CDW Data Request Form may be submitted prior to IRB approval; however, your study must be approved by the IRB – and all data requested from the CDW must be reflected in the approved protocol – before the CDW team can begin to work on your request. CDW analysts will confirm that all variables requested are reflected in the approved IRB before starting the project. Researchers may request more variables in their IRB than they request from the CDW. For example, a research team may conduct a chart review itself and only require the CDW to pull 5 of the 20 variables listed in their full IRB protocol. Teams should be specific about which variables they are the requesting from the CDW and only include variables they are requesting from the CDW in their data request form. Discrepancies between IRB protocols and CDW requests will be returned to investigators and will cause delay in processing CDW requests.
- *It is the responsibility of the investigator to let the CDW know when the study is IRB approved.* Analysts do not monitor initial IRB approval or amendment approvals. The CDW does not being work until notification of IRB approval is received from the investigator.

Step 3. Develop and Submit a CDW Data Request

Data requests may be submitted via the CDW website using the CDW Data Request Form.

Developing a CDW data request

When developing your CDW data request, please be sure to have the answers to the following questions:



Requests for single data extractions vs. recurring data extracts

Investigators need to submit a new CDW Data Request Form for each data set file requested (i.e. each new CDW extraction) – even if the study team previously received a data set from the CDW – to confirm data elements required for the specific timepoint. For example, if an investigator requires recurring data extracts (e.g. quarterly) the investigator is responsible for submitting a new data request form each quarter 30 days before the data set is required. The form should describe the request is for a recurring extraction project and outline the data needed for the specific request/timepoint. Requests for recurring data extracts receive expedited priority.

Billing account information

The requestors must provide account information and the name of an administrative contact in the data request form. See Billing & Invoicing for more details.

Step 4. CDW Reviews the Submitted Request

The CDW aims to review all data requests forms for completeness and feasibility within 5 business days of submission. Revisions and clarifications may be requested. In some cases (for example, when the request provides too few details for adequate evaluation), the CDW will require a consultation; in this case, investigators may incur charges for analyst time (See Billing & Invoicing). Analysts may engage CDW leadership on a case-by-case basis to determine if a request is infeasible or should not move forward.

Step 5. Investigator Receives Data and Confirms Data Completeness/Accuracy

CDW analysts deliver all data sets using BMC's secure box.com system. CDW analysts grant access to the data set to investigators and their designees. These designees must be listed on the approved IRB protocol. The CDW does not email password-protected excel files to investigators.

Authorization to access data

Investigators are responsible for monitoring data access and data security per applicable regulations and guidance. Authorization to access data provided by the CDW is determined by BUMC IRB. Investigators are expected to monitor data access per the study's IRB protocol upon receiving the data set from the CDW. Investigators are responsible for managing IRB amendments to add study team members.

Data extract questions or modification requests

CDW analysts are available to respond to questions about your data set – or make routine modifications – within two weeks of supplying your data set. Questions and modification requests submitted within the two-week deadline will receive expedited priority.

Requests received beyond the two week period require the investigator to complete a new CDW Data Request Form, and the request will be added to the existing data request queue.

Analyst time making routine modifications is subject to fees. In the event the investigator requests modifications that differ substantially from what was previously discussed during the initial request process, the analyst will provide an updated cost estimate. Examples of substantial modifications include new project components, additional variables, different dates, and complex changes to codes or variable definitions.

Step 6. Investigator Receives Final Invoice

Investigators receive notification of final invoiced payment the month following final data set delivery. See Billing & Invoicing.

Once a request has been closed out (two weeks after data delivery if no modifications are requested), all subsequent questions, even if about the existing data set, require entry of a new request form in the queue (Step 3 above) and new invoices will be generated.

Process to Obtain Approval to Access Data from Different Sources

Investigators may need to acquire special permission to access certain data. The table below outlines additional permissions (beyond IRB approval or exemption) that may be required before the CDW may begin to fulfill a data request.

Data Source	Special Approval to Access Data Required?	How to Get Approvals
BMC EPIC Clinical Data	No	N/A
BMC Legacy Systems	No	N/A
CHC OCHIN EHR Data	Yes	Access to OCHIN data requires signed approval from each individual CHC's CEO. For help facilitating the approval process, please contact Allyson Richmond, Program Manager, Boston HealthNet, Allyson.Richmond@bmc.org

Timeline to Receive Data

Wait times for data delivery depend on a variety of factors, including the current request queue and scope of the project. Wait times fluctuate throughout the year. It is the responsibility of the research team to allow ample time for IRB approval and completion of the CDW data request. Research teams should plan on at least a 4-6 week wait between IRB approval and receipt of the data.

Expected wait time by request type provided below.

Data Request Type		CDW Aims to Provide Data to Investigators…	Note
Simple counts for research proposal preparation		14 business days from request submission	CDW provides three free hours of count request services with the expectation that funding for further CDW services are included in award budgets.
Research studies requiring IRB approval/exemption	Initial request/ extraction	4-6 weeks after data request submission/IRB approval	The CDW timeline does not begin until the IRB application is approved.

	Modifications to a data set that was provided by the CDW within the last two weeks	Modification requests submitted within two weeks of receiving your data set receive priority in the queue.	Requests received beyond the two week review period or requests that reflect substantial changes to what was previously discussed with the analyst require the investigator to complete a new CDW Data Request Form, and the request will be added to the existing CDW data request queue.
	Subsequent/ recurring extractions for an ongoing study	30 days from request submission	This timeline assumes the investigator is requesting the same data as the previous extraction and there are no changes to project scope.

Request Prioritization Policies

During high volume periods, the CDW team may have to prioritize requests to maximize workflow efficiency and address BMC's institutional needs.

Projects

Requests are prioritized based on several factors to maximize efficiency and keep the queue progressing. Requests may be prioritized by request type, complexity, size, scope, and which data streams need to be accessed. Questions and modification requests submitted within two weeks of receiving a data set from the CDW receive priority in the queue. See Data Extract Questions or Modification Requests. Requests for recurring extracts also receive expedited priority in the queue. See Requests for Single Data Extractions vs. Recurring Data Extractions.

Investigators

As a BMC Research Administrative Core, the CDW is predominately a BUMC resource. BUMC investigators (defined as investigators from BMC, BUSM, BUSPH, and BUSDM) have first priority for data access. The CDW also provides services to investigators from the Boston University Charles River Campus. At present, the CDW does not accept requests from investigators outside of BUMC or BU.

Research Conducted by Trainees (Students, Residents, and Fellows)

The CDW supports research among Boston University medical students, and residents and fellows from all BUSM/BMC departments. The CDW monitors all services provided to trainees, by department, to ensure equitable distribution of resources.

Students

BU students from the School of Medicine, School of Public Health, and School of Dental Medicine are required to have a faculty-level mentor who oversees the CDW data request process and is ultimately responsible for meeting the investigator expectations outlined in this manual. Students are subject to the hourly fees for CDW services provided. Students should work with that faculty mentor and advisory systems within BU to determine resources to support the research.

Residents and Fellows

Residents and fellows conducting research that requires CDW data are considered members of their respective departments and are subject to the hourly fees for CDW services provided. Residents and fellows are required to have a faculty mentor who oversees the CDW data request process and is ultimately responsible for meeting the investigator expectations outlined in this manual. Residents and fellows should speak with their mentors and departments regarding resources to support their CDW data request.

If you are a student, resident, or fellow conducting research and are concerned about CDW fees, please reach out to the CDW at cdw@bmc.org to schedule a consultation.

Other Avenues to Data Access: TriNextX

TriNetX is a self-service platform that allows BUMC researchers to identify cohorts, participate in sponsored trials, understand therapy and disease landscapes, and conduct analyses without statistical software. A subset of data from Epic and legacy EHR systems (Logician/Centricity) is transformed into a de-identified limited dataset and uploaded to TriNetX every 1-2 months. Dating back to 2000, the data in TriNetX include diagnoses, medications, laboratory results, encounters/visits, vital signs, and demographics.

TriNetX can be used to:

- Determine eligible patients based on your study eligibility criteria to pursue a grant or identify patients for recruitment.
- Participate in sponsored trials, and receive trial requests from partners and sponsors.
- Analyze risk factors, compare treatments.
- Obtain a risk ratio and Kaplan-Meier analysis without statistical software.

BUMC researchers with Human Subjects research training and an active BMC account may use TriNetX. Contact Nicholas Trombley, Research Coordinator, BU CTSI, at Nicholas.Trombley@bmc.org for more information or to request access.

BILLING & INVOICING

Billing Information

BMC has invested in the creation and maintenance of the CDW. Funds collected for CDW services are used to cover services for investigators to continue accessing these data for research purposes. Account information must be provided by the requestor via the request form.

Cost Structure

The fee structure of the CDW is maintained as a BMC Research Administration Core. Currently, all services are billed at the same rate established by the Core. The fee structure is reviewed annually. Currently, the CDW charges $75 per hour for all projects.

Billable Services

All project-specific activities and services beyond the initial hour of free consultation are billable. This includes consultations to further refine the work, develop variable definitions, link data sets, and finalize data format. It also includes data extraction, request follow-ups, and modifications to existing requests.

Services for Which Fees are Waived

A one-hour initial consultation is provided to all studies free of charge.

If the project is a simple counts request to prepare a research proposal, the CDW provides three free hours of services to investigators with the expectation that funding for further CDW services are included in award budgets. Count requests requiring more than three hours of analyst time are subject to the $75 per hour fee.

CDW collaborations with Departments may result in generalizable processes that benefit the CDW and BMC as an institution. In these cases, the CDW may waive a portion of its fees. Such projects are discussed with CDW leadership and/or the Advisory Committee.

Billing Policies

Account information must be provided by the requestor via the data request form. Data sets will not be provided to investigators until the CDW confirms account information has been supplied.

The investigator and administrative contact receive one invoice per delivered data set. For example, a project with a single data extraction will receive one invoice after delivery of the data set. A project with quarterly data extractions will receive one invoice per quarter after the delivery of each quarterly data extract. All analyst hours attributed to the supplied data set are included on the invoice.

Investigators are given two weeks from the date of data set delivery to review the data and request modifications (See Data Extract Questions or Modification Requests). The invoice is generated after this two-week review period or, if modifications are requested, after the modifications are made.

The CDW Team bills in cooperation with BMC Research Administration, which processes all journal entries the first week of the month following invoice delivery.

CDW TEAM OPERATIONS & MANAGEMENT

CDW Analyst Standard Operating Procedures and Coding Conventions

This section provides an overview of standard operating procedures for CDW Analysts and staff to maintain the integrity, quality, availability, confidentiality, and distribution of CDW data.

Opening New Data Requests

CDW analysts aim to review submitted CDW data requests within 5 business days of submission. This initial review will occur even without IRB approval.

CDW Project Tracking System

CDW analysts log all projects into the Project Tracking System in real time as they are submitted via the request form. CDW analysts enter hours worked on specific projects daily; these hours are reviewed at the end of each week.

Coding Conventions

CDW analysts are in the process of developing an internal guide to standardize coding conventions. While these conventions are being developed, analysts will follow a series of general guidelines to facilitate code sharing within the CDW and across other BMC teams. General guidelines include:

- Table naming conventions: When creating new data tables from raw CDW data, analysts should include the PI's last name, IRB number, and a descriptor for the table in the table name (pilast_irbnum_cohort).
- Queries for specific projects should be annotated such that the project's logic can be followed by another analyst.
- Queries to be included in a shareable "code library" should contain the following header:

```
/*****************************************************************************************************************/
/*****Created:          dd/mm/yyyy                                                                          *****/
/*****Created by:       Analyst name                                                                        *****/
/*****Purpose:          Description of the query                                                            *****/
/*****IRB referenced:   IRB if applicable                                                                  *****/
/*****Tables used:      any reference tables used                                                           *****/
/*****Version history:  version history of the initial build and initial review                             *****/
/*****************************************************************************************************************/
```

- Queries for the code library are subject to further scrutiny and must be thoroughly commented out throughout.
- Queries for projects that are known refreshes are to be annotated and organized in a way that enables the analyst to refresh the data set and provide the investigator with the exact same extract, unless agreed upon between the analyst and research team.

Ongoing Project Monitoring

All CDW data requests and projects are monitored by CDW leadership to ensure resources are prioritized and distributed fairly. In certain circumstances, the CDW many need to discontinue providing services to an investigator.

CDW analysts will communicate to the investigator when there is a significant change in the cost estimate (±20% deviation) to avoid unexpected charges.

Delivering Data Sets to Investigators

CDW analysts deliver data sets using BMC's secure box.com system. Analysts grant access to the data set to investigators and their designees only if these study team members are specifically named on an approved IRB protocol. Emailed password-protected Excel files will not be provided.

Data set file specifications

Unless discussed and agreed upon by the analyst and research team, format of the data set is at the discretion of the analyst. The CDW team and Advisory Committee regularly monitor common asks of data set delivery and evaluate solutions that are of universal benefit.

Data dictionaries

A study-specific data dictionary accompanies all data extracts. This data dictionary includes a list of all variable names and specifications in the data set.

Analyst's Project Close-Out Procedures

CDW analysts are available to respond to any questions or clarifications within two weeks of supplying the data set to investigators. It is the responsibility of the research team to review the data set upon receipt and get in touch with the analysts if there are any questions, concerns, or necessary modifications. If the research team has questions or requests additional modifications to the data set beyond the two-week review period, the CDW analyst directs the investigator to submit a new data request form and the query is placed into the analysts' general project queue.

Upon completion of a request, the analyst will enter the request completion date within the tracking system and confirm billable hours with the CDW billing coordinator.

Administrative Management

Dashboarding

Dashboards providing a visualization of key CDW indicators are generated monthly. Dashboards are distributed to CDW leadership for continuous self-assessment and to monitor CDW resource allocation.

Training CDW Staff

Prior to the development of a formal training manual, analysts are expected to follow general guidelines outlined in this procedure manual and as communicated by members of the CDW team.

Policies for Amending CDW Scope, Policies, or Procedures

CDW leadership conducts continuous self-assessment and monitors opportunities for program improvements. The Scientific Director and Director of the CDW formally review this manual every 6 months to ensure accurate and updated policies and procedures are reflected.

CDW leadership may implement changes related to CDW scope of services, policies, or procedures in line with resources and demand.

CDW leadership may engage the CDW Advisory Committee in a formal ratification process for major changes, updates, or revisions. Most changes may be approved via a simple email sign-off process to confirm the change, update, or revision.

For complex changes, updates, or revisions that require discussion, the topic is added to a CDW Advisory Committee meeting agenda.

Leadership Team

The CDW Team reports to BMC's Associate Chief Medical Officer for Research and is overseen by BMC's Chief Medical Office.

CDW Advisory Committee

The Committee is charged with assisting CDW leadership in determining the scope of CDW services, assessing whether the CDW is fulfilling its mission to provide high quality data efficiently for research purposes, and identifying opportunities for improvement and growth.

The CDW Advisory Committee convenes on an ad hoc basis to ensure that committee members' time is used efficiently. In lieu of meetings, committee members are consulted electronically as needed.

FY20-21 Advisory Committee Members

Bill G. Adams, MD
Rebecca G. Mishuris, MD, MPH, MS
Tuhina Neogi, MD, PhD
Jake Nudel, MD
Allan J. Walkey, MD, MSc