Cohort in a Box

TABLE OF CONTENTS

INTRODUCTION

I.　　Rationale

Johns Hopkins clinicians provide care to many unique patient populations. Substantial institutional investments have been made to accelerate knowledge and discovery through development of Precision Medicine Centers of Excellence (PMCOEs), which are supported through integration of various data sources in the Precision Medicine Analytics Platform (PMAP). PMCOEs are dependent on clinical cohorts (well-phenotyped patients followed over time with informative clinical, patient-reported, laboratory and imaging data linked to appropriate biospecimens). Up-front attention and guidance to the rationale, design, and practical aspects of cohort development will improve efficiency, quality, and scientific rigor while adhering to HIPAA requirements and establishing best practices in data collection and management.

However, faculty have varied levels of support, skills, and aptitude for cohort implementation. Many faculty members have struggled to assemble prospective data on their patients, often using ad hoc, retrospective reviews of existing medical records. The limited data collection involved is rarely inclusive of comprehensive phenotyping, patient-reported outcomes (e.g., key behavioral, exposure or disease-specific severity scales), or linkage to available biospecimens. Systematic, prospective collection and integration of comprehensive clinical and patient-reported data linked to appropriate biospecimens can provide the critical research infrastructure to greatly improve the quality and level of research investigation possible and provide a pathway to scholarly publications and further funded research.

II.　　Deliverables

Tangible deliverables expected from this initiative:

    a. Provide one-stop, guided steps for clinical cohort organization (targeting but not limited to PMCOEs)
    b. Strengthen faculty interest and expertise in cohort design, research-grade data collection and management, and analysis
    c. Share best practices: Create model paths for research infrastructure development and utilization that faulty can use as a guide to navigate the complexities of cohort development
    d. Increase quality and scope of preliminary data and supporting infrastructure to strengthen PMCOE, NIH, and foundation grant applications
    e. Ensure that PMCOEs have quality data to best leverage PMAP resources
    f. Provide resources for targeted translational investigations
    g. Serve as a patient resource to facilitate participation in clinical trials from various sponsors
    h. Increase collaboration and co-mentoring opportunities

III.　　Terminology

We advocate for the use of "cohort" rather than "registry." Cohort implies systematic observation of a population with a shared health state or exposure; registry implies a case series, which is a weak design and often involves passive collection of conveniently available non-systematic data.

IV.   Overview

Cohort in a Box involves a series of modules that walk through each step in the process of cohort development. These modules may be of varied relevance for different faculty groups, but we recommend that each section be considered with faculty actively evaluating and focusing on the components that are most pertinent to them. Faculty groups can gain hands-on experience utilizing the modules during training sessions and will also be able to access online modules to develop their research protocols.

MODULES

Cohort Development

> This first step in developing a cohort will be to identify the target patient population, the outcomes of interest, and the scientific question that is being explored.

I.   Cohort Rationale

    a. What is the unique patient population or outcome of interest?
    b. Is your outcome a diagnosis, treatment, shared exposure, or other?
    c. What are your key scientific hypotheses?
    d. In support of your scientific aims, what is the practical rationale for developing the cohort?
        i.   Obtain representative data from a unique population
        ii.   Define and identify subgroups with differing prognoses
        iii. Evaluate possible risk factors
        iv.   Define incident endpoints or complications
        v.   Establish a temporal sequence
        vi.   Prepare for randomized trials
        vii. Establish a research program

> Understanding the landscape for your developing cohort helps to identify existing patient populations that are being studied at your organization or at other institutions. . Landscaping also helps determine the various locations where patients may be seeking care, which can additionally identify potential collaborators.

II.   Cohort Landscaping

    a. What other single or multi-site cohorts exist in your field?
    b. If applicable, how are those cohorts funded?
    c. How does your proposed cohort differentiate itself from those already in existence?
    d. What is the anticipated number (N) of patients and number of patient encounters annually?
    e. Would your cohort likely be acceptable for inclusion into the larger consortia?

    f.   What are the specific referral and care patterns?

    g.   What subspecialties and clinical sites are involved in the care of this population?

III.   Cohort Design

    a.   Will you use prospective data (data collected going forward) or retrospective (non-concurrent data obtained from older records or a combination of the two? You will need to understand the level of detail available in retrospective data and how these may be integrated into prospective collection.

    b.   What will be your baseline (diagnosis, treatment initiation, index clinic visit, etc.)?

    c.   What is an appropriate comparison group? Sometimes the primary comparison will be internal only (e.g., compare based on cohort phenotyping). Is there another relevant comparison group (e.g., other cancer cases, other GI endoscopic patients)?

    d.   Will enrollment be open (continual recruitment), closed (initial recruitment, then just follow those individuals), or periodic (successive waves of recruitment)?

> It is important to determine at what disease stage patients will enter cohort follow-up, particularly as resources to perform deep phenotyping on all subjects may be limited. Later stage disease patients may have a higher rate of hard clinical endpoints. However, early stage disease patients might be amenable to development of novel biomarkers or disease modifying therapies.

IV.   Participant Recruitment and Retention

    a.   What are your detailed eligibility and exclusion requirements?

    b.   How will you identify participants?

        i.   Will you extract qualifying data from EMRs and contact patients? (EPIC for Scholarship: Slicer Dicer, Precision Medicine Analytics Platform)

        ii.   Will you recruit through advertisement with flyers, patient visits to clinics, or letters?

    c.   Details on the recruitment process and copies of recruitment materials are often required by the Institutional Review Board (IRB). In your recruitment materials, be sure to include:

        i.   Name and description of the study

        ii.   Name of investigator(s) and relevant affiliations

        iii.   Contact phone number to reach the research team

        iv.   Eligibility criteria (can be generic, such as "smokers needed for study," just screen for more in-depth criteria over the phone or in person)

        v.   Time commitment required for the study

        vi.   Location of the study

Sample recruitment flyer is available at the end of this document.

    d.   How will you capture participants that do not enroll into your cohort? It is vital to characterize the basic demographics, disease stage, and treatment parameters of individuals who are not approached or who declined participation in your cohort. You need to be able to state the patient population that your enrolled cohort is represents(e.g., all early stage disease patients receiving care at a tertiary referral clinic).

> Variations in recruitment may be related to specific referral patterns, available research staff for recruitment, or other unmeasured factors that could result in biased selection of cohort participants. For example, Dr. A recruits all of his patients with disease X into the cohort during his Thursday clinic. However Dr. B, who does Wednesday clinic, does not actively engage patients to become involved in the cohort. Another example would be when higher income or employed patients do not participate because they are unwilling to commit the time required for additional study procedures. In contrast, heavily engaged patients may be very proactive in seeking study participation and may not be reflective of all patients who have the disease.

    e. What strategies will you employ to retain your participants (e.g., telephone follow-up, My Chart contact, incentives, parking or meal vouchers, newsletter or dissemination of study results)?

Data Considerations

I.   Data Sources

    a. EMR data (Hopkins EPIC, CRISP, etc.)
    b. Clinic visits or research study visits or both
        i.   Interviews
        ii.   Clinical examination
        iii. Laboratory data (routine or study specific)
        iv. Physiologic tests
        v.   Imaging
    c. Administrative data (e.g., diagnostic and discharge codes, insurance, Medicare/Medicaid linkages)
    d. Registry linkage (e.g., cancer, transplant)
    e. My Chart, telephone, or email contact/follow-up
    f. Digital and mobile monitoring, wearables

> For each data element, you will need to consider what data source(s) may be possible (required) and how they will be obtained.

II.   Data Elements

    a. Develop an exhaustive list of potential data elements, often grouped by demographic, socioeconomic, behavioral, clinical, physical exam, routine or special labs, pathology, imaging, and other relevant categories
    b. Will you be using disease-specific questionnaires?
        i.   Is there a standard measure used in your field (e.g., St. George's COPD scale)?

> When possible, use standardized instruments with widespread application (e.g., AUDIT for hazardous adult alcohol use; see PROMIS or PhenX for comprehensive list).

      ii.   If you developed your own measure, how did you validate it in comparison to the accepted "gold standard" measure?

> **Illustrative example**: You may be interested in measuring patient activation as a determinant of response to treatment. Consider a scenario where there are three different survey instruments that attempt to measure patient activation; how would you choose the appropriate measure? Which one is used most commonly in clinical or research settings? Which has the best performance characteristics? Which one is most appropriate for your study population or disease of interest?
>
> This will be based on prior published and possibly unpublished data. If there is a clear gold standard instrument that is widely used, use that instrument. Brevity is also preferred. Is the 12 item instrument as reliable and valid as the 22 item? Additionally, are there licensing fees involved?

c. Is the mode of delivery appropriate?
      i.   Should it be electronic / digital (e.g., If the instrument was developed as an interviewer-administered survey, is there data to support its' translation to computer-administered or My Chart survey)?
      ii.   Is the language used in the instrument appropriate for the reading comprehension level of your particular study population?
      iii.   What type of training is required to administer the instrument, and who has the expertise to execute the training (e.g., healthy literacy assessment is suggested to be administered by specially trained individuals)
d. How are you performing longitudinal data collection? Will you just be moving forward prospectively or also going back to older records. How will the data capture, abstraction or review process for these two sources of data be different?

> Especially if going back to retrospective data collection, you will need to pay attention to temporal changes in how the data may have been collected (e.g., Beck Depression inventory has been asked for three years, followed by a switch to the CES-D). For many common instruments, these data may have already been mapped in algorithms for combining data, which will allow data integration for continuity purposes. PROsetta Stone links instruments to PROMIS metrics.
>
> Another example would be evolving laboratory methods over time (e.g., lower limits of detection of HIV RNA, such as a decline in the sensitivity of detection from 1,000 to 400 to 200 to <50 copies/mL). For laboratory results, in addition to the assay value, you need to get the specifics of each individual assay performed along with its specific dynamic range, including the limit of detection. The assay information should be specific data elements collected along with the actual value. Another important consideration is retention of the raw laboratory data. For example, if an assay is performed in triplicate and only the mean is incorporated into the database, this will not allow assessment of coefficients of variation or identification of outliers. It's preferable to retain all three values which allows calculation of the mean for subsequent analysis.

e. What is the file output for each data element particularly with regards to data coming from multiple sources (e.g., CSV files generated directly from the flow cytometer machine)?

f. Will the data be continuous, ordinal, or categorical?

> Plan for range and edit checks now. Sample output should be obtained from each contributing laboratory or data source to ensure that the appropriate fields are created in the database.

> There will likely be variables created (called derived variables) including combinations of these data elements into specific indices or categories to create composite variables (e.g., metabolic syndrome, Childs-Pugh stage). Ensure that you have included all components of these indices in a format that allows creation of index variables.

g. What is the acceptable range of values?

III. Data Collection

a. What type of tool will you use to collect your research data?
    i. Will it be an electronic survey administered as part of the study visit (RedCap, Qualtrics, Survey Monkey, QDS, etc.)?
    ii. Will EPIC smart forms be developed with data captured as part of the clinical visit?

> Initial meetings should identify standardized measures, skip patterns, and other items that should be included in your data collection tool. Additionally, you will need to address how to establish IDs and how to connect different surveys using the study IDs. You will need to create a master document with the above items along with detailed instructions Standard Operating Procedures (SOP) for survey administrators.

    iii. Will you use paper and pencil collection (and if so what about plans for data entry and verification?)
b. Where will the data collection happen?
    i. Will you send an electronic survey link directly to the patient (and will this require e-signature on consent form)?
    ii. Will you collect data in-person?
        i.   Will it occur in the clinic?
        ii. Is it a pre-visit, post-visit, or separate study visit from the normal clinic visit?
        iii. How will you handle patients that live farther away? Or that are employed?
    iii. Will you use a Clinical Research Unit (CRU)? CRUs provide sites for comprehensive data collection with personnel that can perform evaluations, administer medications, perform phlebotomy services, etc. CRUs require a separate approval process and have implemented cost recovery processes.
    iv. Will you conduct phone interviews? If so you will need scripts and rules for contacting and leaving messages on individuals bones as well as standardized data collection tools
c. What additional information will be needed and how will it be collected?

      i.   Do you need additional contact information to track or follow-up with participants or will the information from medical record suffice? Locator information collected outside the EMR needs to be handled with extreme safeguards to protect patient privacy.

      ii.   Do you need a payment form to document payments made to participants?

Are you using multiple forms? If so, you will need a checklist form to ensure that all the appropriate study evaluations and their respective forms have been completed. The checklist can also include any eligibility criteria that ensures only eligible patients are included in the study or undergo specific procedures.

IV.   Database Creation and Maintenance

a. Identify appropriate software to use for data collection, data management, and data analysis. These may be different software applications used in conjunction (e.g., RedCap for data collection, Stata or R for data analysis)

b. If resources do not permit hiring a data analyst, consider working with graduate students to develop a database

c. Identify integration issues, such as merging data from EMRs to data collected from the study visits, and develop standardized formats

d. Document the specs of the database

e. Identify server location(s) to store the database that is secure, HIPAA compliant, and has acceptable backup or restoration procedures (See these data archiving and storage resources)

f. Consider using a Secure Analytic Framework Environment (SAFE) desktop from Hopkins to share and analyze sensitive data

g. Importantly, identify the type of reports that will need to be retrieved from the database. Preferably, work with a statistician to create data collection and data management plans to reflect future publications and presentations. Check out on-campus statistical consulting cores such as ICTR and BEAD

h. Decide whether web-based access to reports (such as a dashboard) will be useful, then identify features that need to be built to support such access

i. Develop processes to ensure data safety and controlled accessibility (passwords, encryption, use of laptop, minimize use of external drives, etc.)

V.   Biospecimen Collection & Repository and Laboratory Testing

a. Identify samples to be collected
      i.   What samples will you collect?
      ii.   How many samples will be collected?
      iii.   How frequently will you collect samples?
      iv.   Where and how will the samples be collected?
      v.   Who collects the samples?
      vi.   How will the samples be stored, processed, and transported if necessary

> The specific type of samples collected will depend on the planned assays and well as ensuring appropriate sample types for future novel testing.

b. For laboratory testing, will the testing be carried out as part of routine clinical care, or will it be done solely for the study?
    i. Will you use an internal lab?
    ii. Will you use a lab at another institution?
    iii. Will you use outside vendors such as Quest or LabCorp?
c. Identify the type of storage required for each sample type
    i. At what temperature should the samples be stored?
    ii. Where will you store each sample type?
    iii. How much are the initial processing fees and annual maintenance fees?
d. Establish a point of contact for each lab or repository site, including for emergencies and after-hours
e. Consult with your lab(s) regarding the type of tubes to use to collect samples, the preservative requirements, the storage instructions between collection and delivery to lab, and the drop-off/pick-up information
f. Establish guidelines for the sample labelling process and how to correct errors
g. Establish documentation regarding collection, transportation (may need a form), storage, retrieval, and obtaining analysis results. OpenSpecimen is encouraged in the Hopkins environment; other software such as Freezerworks are also commonly used)
h. If samples are to be divided for multiple uses, declare clear guidelines and priorities for aliquots, labeling, and storage after the sample reaches the lab; you will need documentation for date, time, and personnel completing those procedures, as well as thorough details of where the sample is stored (freezer, shelf, and/or box number, etc.)
i. Establish secure ways of receiving results from labs if tests are performed immediately
j. For a repository, establish guidelines to identify samples (e.g., create plans for different combinations of categories that may be used later for repository pulls)

Human Resources

I. Team Training, Structure, and Roles

a. Based on the needs of the study, identify necessary personnel
    i. Do you need a study manager? The study manager is responsible for the administrative, regulatory, and technical details. This person will also provide oversight of other team members and will function as a liaison between different key entities (investigators, staff, laboratories, data management, regulatory bodies, etc.)
    ii. Do you need field staff/data collectors? Do they need to have a clinical background or can non-clinical staff with appropriate training suffice?
    iii. Do you require a data manager or statistician?
b. Estimate the level of effort for each personnel category

> In many instances, a single study may not need full-time employment for some positions, and substantial cost savings may be realized through sharing personnel with other studies. Be sure to clarify the specific time commitment (e.g., number of hours, specific days), study roles, office space, and reporting structure for shared staff.

c. How will your recruit for staff positions?

       i.   Internal vs. external
     ii.   Working with your departmental HR to post positions

> Remember, hiring generally takes one to three months to complete, so factor this timeline into your project implementation plans.

d. Initial training should include:
   i. HIPAA
   ii. Human Subjects
   iii. Conflict of Interest
   iv. Project-specific trainings
   v. Handling unanticipated situations and subsequent reporting requirements
   vi. Accurate and appropriate completion of documentation
   vii. Data collection procedures and use of databases
   viii. Data safety, protection of privacy, confidentiality
   ix. Problem solving for likely scenarios
e. Will staff need retraining?

> Periodic retraining is appropriate if your study has multiple interviewers/examiners. Be sure to have systems in place to identify any discrepancies in approach to data collection early and to resolve these discrepancies.

Regulations and Monitoring

> These regulatory processes aim to protect the rights, safety, and welfare of the study participants from the time of identification of potential participants to completion of the study.

I.   IRB and Regulatory Issues

a. Do you require written informed consent, or is verbal assent sufficient?
   i. Develop informed consent form and process to obtain consent, including who has the authority to obtain consent
   ii. Be sure that your consent form is professional yet at an appropriate reading level (capable of being understood by a lay person unfamiliar with medical or technical terminology)
   iii. Follow templates provided by the IRB
b. Prospective Reimbursement Analysis (PRA): Clinical Research and Support Services (CRSS) will determine the potential charges to be considered standard of care or research, then CRSS will upload the resulting PRA to the study's eIRB portal once the IRB application is submitted
c. The CRSS will also complete the Insurance and Research Participant Financial Responsibility Information Sheet, which provides cost information to the participant. CRSS will provide a stamped copy to the investigator
d. Documentation of required and trainings/certification of investigator(s) and staff
   i. Human Subjects Research
   ii. Clinical Research and Billing Orientation (CRBO) Training to understand what aspects of a visit are billed as research

      iii.   Clinical Research Management System (CRMS) Training to learn how to use a portal to track patients in research studies, a system which can be used to generate reports annual regulatory approvals

    e. Determine if clinicaltrials.gov registration is required and ensure compliance

II.   Quality Assurance and Quality Control

a. Use checklists to ensure all study procedures are completed and to track if items were not completed or if there are consent variations (e.g., no genetics consent)
b. Consider a complete review of the first 20 participants recruited, followed by conducting random checks on later participants to ensure high-quality completion of all documentation
c. Develop and implement a system to identify protocol errors
d. Document changes in processes and when new processes are implemented, ensuring training and adherence by staff

Other Considerations

> A well-developed, comprehensive protocol can be used by the research team to gain a thorough understanding of the research project.

I.   Developing Protocols

a. What is the background and rationale for the project?
b. What is the research question? What are the relevant hypotheses?
c. What are the objectives or specific aims of the study?
d. What is the study design?
e. What is the study population and sample size?
f. What are the exposures and outcomes, including number of exposures and details of processes at each encounter?
g. What are the methods for data collection?
h. What are the analytic methods to be employed?
i. What is the timeline for the project?
j. What are the specific Human Subjects considerations?

II.   Budgeting

a. Identify direct costs
    a. Salaries
    b. Equipment
    c. Supplies
    d. Laboratory testing
    e. Biorepository storage costs
    f. Participant reimbursement
    g. Conferences and travel
    h. Software and hardware expenses

b. Use your institution's established F&A (indirect) costs. If this funder will not pay agreed-upon rates, need to get approval from your administrator

c. Review your study timeline and make sure the funds requested adequately cover each year of the project

d. If unsure of where to start, review similar studies for a ballpark budget

e. Make sure the funds you request are allocable, allowable, and reasonable

f. When in doubt, consult your budget analyst for more details

III. Communications

a. Establish procedures for PI(s) to regularly meet and communicate with study coordinator regarding updates on recruitment, study procedures, challenges, and personnel issues

b. Study coordinator and study staff should establish periodic team meetings and clear methods for communication of general study issues, adverse events, and HR issues

c. PI(s) and study coordinator should meet with budget analyst at least quarterly (consider monthly for more complex budgeting) to review expenditures and budget projections

d. Study investigators should meet periodically to discuss study progress and share preliminary data

e. Clearly delineate issues of leadership on particular analyses or manuscripts (ideally, authorship guidelines and expectations should be discussed in advance)

f. Consider a dissemination plan for study results, such as establishing a website, blog, or working with your press office

IV. Collaborations

a. Establish a procedure for external investigators to develop a collaboration
   i. You may include a formal discussion with the PI(s) to develop a concept sheet and have it reviewed by an executive committee
   ii. Consider project agreement (data use agreement, materials transfer agreement) which establish guidelines for data security, authorship and acknowledgement, and financial responsibilities for collaborative studies

b. Ensure compliance with NIH data sharing plan as appropriate

SAMPLE RECRUITMENT MATERIAL

For more information

Interested in learning more?

Contact one of our Research Assistants:

Ladora Cromwell

Phone: 443-287-9042

Email: lcromwe2@jhu.edu

OR

Sara Boynton

Phone: 443-287-9051

Email: sboynto3@jhmi.edu

They will tell you more about the study, determine if you or your child is eligible, and help arrange for a study visit!

Principal Investigator

Tammy M. Brady, MD, PhD
Johns Hopkins University
School of Medicine
200 North Wolfe St.
Baltimore, MD 21287

410-955-2467

IRB00065277

Johns Hopkins University

Electronic Media Powering positive health changes in youth

EMPower

A Research Study funded by:

The National Institutes of Health/National Heart, Lung and Blood Institute

Source: Practical Skills in Conducting Research in Clinical Epidemiology and Investigation (Jacbosn, McKay 2018)