

Tips for Completing a PaTH Data Request

1. **Specify which institutions you hope to engage in your study.** PaTH is a collaboration between the University of Pittsburgh, Penn State University, Temple University, Johns Hopkins University, Geisinger Health System and the University of Utah. At least 2 institutions must be involved for a PaTH-affiliated study.
2. **Clearly describe the eligibility criteria for your sample.** For example, investigators often specify age, sex, ICD-9 codes for relevant diagnoses, and the time-frame of interest. If the eligibility requirements include multiple criteria, please group the inclusion criteria with AND statements and explain the rules for exclusion if any. If you are only interested in data from specific sources, that should be noted as well. Please keep in mind that PaTH has a small number of predefined populations (e.g., “Loyalty” cohort) with well-characterized data that could be drawn upon for your request.

EXAMPLE: Sample eligibility

Inclusion criteria: (“AND”: Must meet *all* criteria)

- Age \geq 18 years old
- At least one outpatient visit in a primary care clinic between 1/1/2012 – 12/31/2014 with associated ICD-9-CM code (primary or secondary diagnosis) for asthma (493.xx)

Exclusion criteria: (“OR”: May be excluded if meets *any* criteria)

- Any encounter (visit or admission) with ICD-9-CM code for cystic fibrosis (277.0x) (primary or secondary diagnosis)

3. **Indicate the data you need.** Be specific. For example, refer to ICD-9 codes and PaTH common data model (CDM) elements (*refer to attached data dictionary*). As noted above, preparatory-to-research data often are displayed as a table describing the eligible sample and the prevalence of the primary outcome and one or two key predictor(s). At times, continuous data such as average blood pressure or weight may be included; please note that such calculations complicate the data request and may require more time to process. If the data needed include elements with a large range of data that you would like categorized, please define the ranges for these categories (e.g., Age per the table below; Blood pressure SBP \geq 140 vs. <140, DBP \geq 90 vs. <90). Be sure to specifically state the relevant time-frame for your data table and which PaTH sites should be included.

EXAMPLE: Please provide data showing the sample’s distribution according to age & cardiovascular history.

1. By AGE CATEGORY: 18 to <30 years; 30 to <65 years; 65+ years
2. By history of myocardial infarction—defined as any history in searchable record of ICD9 410.xx (acute MI) or 412.xx (old MI)
3. By known cerebrovascular disease—defined as any history in searchable record of ICD9 430-438

Table Shell:

	PaTH Site 1	PaTH Site 2
Patients in cohort (from Part A)—n		
Age category—n (%)		
18 to <30 years		
30 to <65 years		
65+ years		
History of myocardial infarction—n (%)		
History of cerebrovascular disease—n (%)		

4. **Generate table shells to be filled in with the data.** For each PaTH CDM variables, include a reference to their CDM table and column names.
5. **Formally submit your data request to the PaTH technical teams.** While we try to ensure that the requests are clear before submission, sometimes questions arise. If so, your PaTH technical liaison team may ask you to help clarify the request. Please use the “PaTH Data Pull Request” form.

Ideas for ensuring that your request is feasible: Be sure that your data request includes each element of the classic “epidemiologic triad” – person (i.e., eligibility criteria), place (institutional location), and time (e.g., dates or visits). Specific tips for these elements are included in the table below.

<i>Do this:</i>	<i>Avoid this:</i>
Person	
Specify discrete data that are stored in the EHR such as numerical fields (e.g., blood pressure, height, weight) or categorical values (e.g., ICD-9 or CPT codes that are present/absent; data entered from a list or checklist; or data entered with a “lookup” field, where a typist must choose between discrete options)	Data stored as a PDF, text document or image. In many cases, you may not be sure, and your PaTH liaison team can help (e.g., Ejection fraction sounds numeric, but may not be stored as a stand-alone numeric fields, but rather, as text).
Be specific in your request (e.g., name ICD-9 codes, CPT codes, refer to specific PaTH CDM elements)	Vague concepts, such as “heart conditions” or “immune-compromised” Physiological criteria that depend on physical and lab findings Qualitative names for procedures (“abdominal surgery”)
Be willing to modify your request to make it more feasible. We’ve found that what sounds simple or complex to clinical researchers is often not well aligned with what the technical teams can accomplish, and that simple adjustments can often preserve the larger scientific goals while greatly reducing technical complexity.	Complex criteria. For example, avoid the use of modifiers (adjectives or adjectival clauses) which contain ordinal or temporal logic. You can ask if a patient has ever had a Hemoglobin A1c done, but avoid asking anything specific about the third one he had done. Likewise you can ask if a patient has gotten a diagnosis of coronary artery disease as well as a prescription for beta blockers but avoid asking if any such prescription happened within 90 days of any such diagnosis.
Place	
Specify the locations as much as possible, e.g., admissions, identity of clinics, outpatient procedures...	Vague descriptions (“outpatient visit” – is same-day surgery an “outpatient visit”?)
Time	
Provide calendar time intervals (dates) that are consistent for everyone in the sample	Data requested within a varying (patient-specific) time-frame (e.g., after the date of a specified visit that differs for different individuals, such as as within “n” number of visits
General	
Specify objective numeric limits for laboratory criteria	Qualitative lab criteria (“hypercalcemia”)
For subjective outcomes, specify measured outcomes, e.g., patient-reported questionnaires	Outcomes that depend on verbal history (which are text)
List specific medications, by generic AND trade names. Consider use of pharmaceutical class for queries involving medications	Medications of interest identified by their therapeutic intention
	Discrete data found in multiple fields of the source system. For example, procedure codes may be found in billing data in one system, and in visit codes, in another. Billing data may not be available at all, and the visit coding may have to be “translated” from one coding scheme (e.g., CPT) to another (E.g., SNOMED). Your liaison team can help you determine when data are stored apart

The best way to develop a request that can be processed more quickly: Restrict the data elements in your request to ones that are already standardized in the PaTH common data model (CDM) or in the data set for one of PaTH’s condition-specific cohorts.

Are you interested in starting a data request? Contact us at PaTHCDRN@pitt.edu

Please let your PaTH liaison team know if you have concerns or questions. We’re continually looking to improve this process and would appreciate feedback.