WashU Medicine

Department of Otolaryngology-Head and Neck Surgery

R25 Predoctoral Mentored Research Training Program

APPLICATION DATES

Application due date is December 31, 2026 for a July 1, 2027 start date
Award Notice Date for applications received by December 31: late February/early March
Start Date: July 1 or August 1, 2027 *(Date to be agreed upon, subject to NIH guidelines)*

APPLICATION PROCEDURE

A COMPLETE PREDOCTORAL APPLICATION CONSISTS OF THE FOLLOWING DOCUMENTS:

- ☐ **Application Cover Sheet** (included below)
- ☐ **Transcripts** from MD degree program. Unofficial copies are acceptable.
- ☐ **One-page Personal Statement**
- ☐ **Five-page Research Proposal** (Helpful resources: *Guidelines for Preparation of Proposals* andPICORI Methodology Standards – included below.)
- ☐ **Mentor Letter of Support**
- ☐ **Reference Letter** (from a previous mentor, supervisor, or individual qualified to speak to the applicant's strengths and fitness as related to the R25 Predoctoral program)

Please note: WashU MD and MSTP students must obtain a letter of recommendation from Terrance Kummer, MD. Additionally, students currently in their MD training will be required to take a 9-month leave from the MD curriculum if accepted into the R25 program. Both of these require an application to the year-long research program (YRP) to be filed with the Office of Medical Student Research and Scholarship. Please submit your YRP application early-at least 6 weeks before needing your letter of recommendation from Dr. Kummer.

SUBMISSION

Please send the application documents to bkatie@wustl.edu

For more information contact:
Katie Buffington, Program Administrator
bkatie@wustl.edu
PH: 314-362-7511

WashU Medicine

Predoctoral Mentored Research Training Program

Otolaryngology-Head & Neck Surgery

Application Cover Sheet

Applicant Information

_____ _____ _____

Last Name *First Name* *Middle Name*

Permanent Mailing Address: _____

Street Address

_____ _____ _____

City *State* *Zip Code*

Primary Phone: _____ Evening Phone: _____

Other Phone: _____ E-mail Address: _____

Date of Birth (mm/dd/yyyy): _____ Country of Birth: _____

State of Birth: _____ Social Security Number _____

Gender: ☐ Female ☐ Male ☐ Intentionally Withheld

Citizenship:

☐ US Citizen ☐ Non-citizen National ☐ Other: _____

☐ Permanent Resident / Non US Citizen with Permanent US Visa

Resident Card Number: _____ Effective Date: _____

Race (check one or more items as appropriate):

☐ American Indian / Alaska Native ☐ Asian ☐ Black / African American

☐ Native Hawaiian or Other Pacific Islander ☐ White ☐ Intentionally Withheld

Ethnicity: ☐ Hispanic or Latino ☐ Not Hispanic or Latino ☐ Intentionally Withheld

Are you from a disadvantaged background? ☐ Yes ☐ No ☐ Intentionally Withheld

Do you have a disability? ☐ Yes ☐ No ☐ Intentionally Withheld

If Yes, which of the following categories describe your disability(ies)?

☐ Hearing ☐ Visual ☐ Mobility / Orthopedic Impairment ☐ Other: _____

Professional Position/Institutional Information

Academic Rank:

☐ Medical Student - What program year will you complete by June? _____

Institution: _____

Education

List all awarded degrees

Degree: ☐ BS ☐ BA Undergraduate GPA _____

☐ Master's _____ MCAT Scores:

☐ Other*

*If Other, Specify Degree: _____

Degree Awarded Date: _____

Institution Name: _____

Have you received previous support from an Institutional NIH training grant? ☐ Yes ☐ No

If yes, indicate the dates of support and the title of the grant or grant number:

_____

Area of interest for which you are applying:

☐ Basic Research
☐ Clinical Research
☐ Translational Research
☐ Behavioral Research
☐ Other: _____

Reference 1

Last Name: _____ First Name: _____ MI: _____

Title: _____ Institution: _____

Phone: _____ E-mail Address: _____

Mentor

Last Name: _____ First Name: _____ MI: _____

Title: _____ Institution: _____

Phone: _____ E-mail Address: _____

A complete submission must include:

- This completed **Application Cover Sheet**
- Transcripts from MD degree program
- **One-page Personal Statement**
- **Five-page Research Proposal** (Helpful resources: *Guidelines for Preparation of Proposals* and *PICORI Methodology Standards* - included below.)
- **Reference Letter** (previous mentors, supervisors, or individuals qualified to speak to the applicant's strengths and fitness as related to the R25 Predoctoral program)
- **Mentor Letter of Support**

I certify that all information provided is complete and accurate to the best of my knowledge.

Applicant Signature: _____ **Date:** _____

Applications and reference letters should be submitted in PDF format to bkatie@wustl.edu and should include "Predoctoral Application: [Applicant Name]" in the subject line.

WashU Medicine

Department of Otolaryngology-Head and Neck Surgery

Predoctoral Mentored Research Training Program

Guidelines for the Preparation of Proposals

After consultation with the proposed mentor, the student should prepare a single-spaced proposal that includes the following sections:

I. Title of Project

II. Name of Trainee

III. Name of Proposed Faculty Mentor and Department

IV. Specific Aims (1 page)

State concisely the goals of the proposed research and summarize the expected outcome(s), including the impact that the results of the proposed research will exert on the research field(s) involved. List succinctly the specific objectives of the research proposed (e.g., to test a stated hypothesis, create a novel design, solve a specific problem, challenge an existing paradigm or clinical practice, address a critical barrier to progress in the field, or develop new technology).

V. Research Strategy (3.5 pages)

a. Significance. Explain the importance of the problem or critical barrier to progress that the proposed project addresses. Describe the scientific premise for the proposed project, including consideration of the strengths and weaknesses of published research or preliminary data crucial to the support of your application. Explain how the proposed project will improve scientific knowledge, technical capability, and/or clinical practice in one or more broad fields.

b. Innovation. Explain how the application challenges and seeks to shift current research or clinical practice paradigms. Describe any novel theoretical concepts, approaches or methodologies, instrumentation or intervention(s) to be developed or used, and any advantage over existing methodologies, instrumentation, or intervention(s). Explain any refinements, improvements, or new applications of theoretical concepts, approaches or methodologies, instrumentation, or interventions.

c. **Approach.** Describe the overall strategy, methodology, and analyses to be used to accomplish the specific aims of the project. Describe the experimental design and methods proposed and how they will achieve robust and unbiased results. Unless addressed separately in the Resource Sharing Plan, include how the data will be collected, analyzed, and interpreted as well as any resource sharing plans as appropriate. Discuss potential problems, alternative strategies, and benchmarks for success anticipated to achieve the aims.

VI. Literature Citations, Progress Report Publication List (include PMCID # if applicable), **and Consortium/Contractual Arrangements** (no page limit)

VII. Secondary Project (1/2 page)

Describe briefly one additional or secondary project that the trainee plans to complete as a supplement to their primary project. The goal of the secondary project is to explore another type of study design and research methodology from the primary project and in this way the trainee will learn additional research skills. In addition, the secondary project should be something the trainee can work during the inevitable stops or breaks in the primary project. The secondary project should be completed with the primary mentor.

Although the written proposal should be the work of the trainee, it is understood that the proposed faculty mentor will provide guidance in its preparation. It should be emphasized that the project should have fundamental scientific merit and some novelty, but should also be designed so that it is feasible to carry out within the anticipated period of the training grant appointment. Please note the new NIH requirements for this section around Rigor and Reproducibility:
http://grants.nih.gov/reproducibility/faqs.htm - 4834

If you need additional information, please contact:
Dr. Jay Piccirillo at (314) 362-8641 or piccirij@wustl.edu

PCORI METHODOLOGY STANDARDS

1: STANDARDS FOR FORMULATING RESEARCH QUESTIONS

RQ-1: Identify gaps in evidence.

Gaps in the evidence identified in current systematic reviews should be used to support the need for a proposed study. If a systematic review is not available, one should be performed using accepted standards in the field (see **SR-1**), or a strong rationale should be presented for proceeding without a systematic review. If the proposed evidence gap is not based on a systematic review, the methods used to review the literature should be explained and justified.

RQ-2: Develop a formal study protocol.

Researchers should develop a formal protocol that provides the plan for conducting the research. The protocol should specify the research objectives, study design, exposures and outcomes, and analytical methods in sufficient detail to support appropriate interpretation and reporting of results. Protocols should be submitted to the appropriate registry (e.g., clinicaltrials.gov), and all amendments and modifications (e.g., changes in analytic strategy, changes in outcomes) should be documented.

RQ-3: Identify specific populations and health decision(s) affected by the research.

To produce information that is meaningful and useful to people when making specific health decisions, research proposals and protocols should describe (1) the specific health decision the research is intended to inform, (2) the specific population(s) for whom the health decision is pertinent, and (3) how study results will inform the health decision.

RQ-4: Identify and assess participant subgroups.

In designing studies, researchers should identify participant subgroups, explain why they are of interest, and specify whether subgroups will be used to test a hypothesis or for exploratory analysis, preferably based on prior data. A study should have adequate precision and power if conclusions specific to these subgroups will be reported.

RQ-5: Select appropriate interventions and comparators.

The interventions and comparators should correspond to the actual healthcare options for patients, providers, and caregivers who would face the healthcare decision. The decision should be of critical importance to the relevant decision makers, and one for which there is a compelling need for additional evidence about the benefits and harms associated with the different options. Researchers should fully describe what the comparators are and why they were selected, describing how the chosen comparators represent appropriate interventions in the context of the relevant causal model (**CI-1**), reduce the potential for biases, and allow direct comparisons. Generally, usual care or nonuse comparator groups should be avoided unless these represent legitimate and coherent clinical options.

RQ-6: Measure outcomes that people representing the population of interest notice and care about.

Identify and include outcomes the population of interest notices and cares about (e.g., survival, functioning, symptoms, health-related quality of life) and that inform an identified health decision. Define outcomes clearly, especially for complex conditions or outcomes that may not have established clinical criteria. Provide information that supports the selection of outcomes as meeting the criteria of "patient centered" and "relevant to decision makers," such as patient and decision-maker input from meetings, surveys, or published studies. Select outcomes that reflect both beneficial and harmful effects, based on input from patient informants and people representative of the population of interest.

2: STANDARDS ASSOCIATED WITH PATIENT CENTEREDNESS

PC-1: Engage people representing the population of interest and other relevant stakeholders in ways that are appropriate and necessary in a given research context.

Include individuals affected by the condition and, as relevant, their surrogates and/or caregivers. Other relevant stakeholders may include, but are not limited to, clinicians, purchasers, payers, industry, hospitals, health systems, policy makers, and training institutions. These stakeholders may be end users of the research or be involved in healthcare decision making.

As applicable, researchers should describe how stakeholders will be identified, recruited, and retained and the research processes in which they will be engaged, Researchers should provide a justification in proposals and study reports if stakeholder engagement is not appropriate in any of these processes.

PC-2: Identify, select, recruit, and retain study participants representative of the spectrum of the population of interest and ensure that data are collected thoroughly and systematically from all study participants.

Research proposals and subsequent study reports should describe the following:

- The plan to ensure representativeness of participants
- How participants are identified, selected, recruited, enrolled, and retained in the study to reduce or address the potential impact of selection bias
- Efforts employed to maximize adherence to agreed-on enrollment practices
- Methods used to ensure unbiased and systematic data collection from all participants

If the population of interest includes people who are more difficult to identify, recruit, and/or retain than other study populations (e.g., individuals historically underrepresented in healthcare research such as those with multiple disease conditions, low literacy, low socioeconomic status, or poor healthcare access, as well as racial and ethnic minority groups and people living in rural areas), then specify plans to address population-specific issues for participant identification, recruitment, and retention.

PC-3: Use patient-reported outcomes when patients or people at risk of a condition are the best source of information for outcomes of interest.

To measure outcomes of interest identified as patient-centered and relevant to decision makers (see **RQ-6**) for which patients or people at risk of a condition are the best source of information, the study should employ patient-reported outcome (PRO) measures and/or standardized questionnaires with appropriate measurement characteristics for the population being studied. In selecting PRO measures for inclusion in a study, researchers, in collaboration with patient and other stakeholder partners, should consider (1) the concept(s) underlying each PRO measure (e.g., symptom or impairment) and how it is meaningful to, and noticed by, patients in the population of interest; (2) how the concept relates to the health decisions the study is designed to inform; (3) how the PRO measure was developed, including how patients were involved in the development; and (4) evidence of measurement properties including content validity, construct validity, reliability, responsiveness to change over time, and score interpretability, including meaningfulness of score changes in the population of interest with consideration of important subgroups as well as the translation process if the measure is to be used in multiple languages. If these measurement properties are not known, a plan for establishing the properties must be provided. Caregiver reports may be appropriate if the patient cannot self-report the outcomes of interest.

PC-4: Support dissemination and implementation of study results.

All study results must be made publicly available. Study objectives and results should be presented in lay language summaries so they are understandable and actionable by as many people as possible. For study results that are appropriate for dissemination and implementation, involve patients and other relevant stakeholders in (1) planning for dissemination from the start of the research study, (2) creating a dissemination plan for the study indicating clinical implications, (3) working with patients or organizations to report results in a manner understandable to and usable by each target audience, and (4) identifying successful strategies for the adoption and distribution of study findings to targeted patient and clinical audiences.

3: STANDARDS FOR DATA INTEGRITY AND RIGOROUS ANALYSES

IR-1: A priori, specify plans for quantitative data analysis that correspond to major aims.
Before analysis is undertaken, researchers should describe the analytic approaches that will be used to address the major research aims. These include definitions of key exposures, outcomes, and covariates. As applicable, study protocols should identify patient subgroups of interest, plans (if any) for how new subgroups of interest will be identified, and how analysis plans may be adapted based on changing needs and scientific advances. Researchers should also specify plans for handling missing data and assessing underlying assumptions, operational definitions, and the robustness of their findings (e.g., sensitivity analyses).

IR-2: Assess data source adequacy.
In selecting data sources and planning for data collection, researchers should ensure the robust capture of exposures or interventions, outcomes, and relevant covariates. Measurement properties of exposures and outcomes should be considered, and properties of important covariates should be taken into account when statistically adjusting for covariates or confounding factors.

IR-3: Describe data linkage plans, if applicable.
For studies involving linkage of patient data from two or more sources (including registries, data networks, and others), describe (1) the data sources and/or the linked data set in terms of its appropriateness, value, and limitations for addressing specific research aims; (2) any additional requirements that may influence successful linkage, such as information needed to match patients, selection of data elements, and definitions used; and (3) the procedures and algorithm(s) employed in matching patients, including the success, limitations, and any validation of the matching algorithm(s).

IR-4: Document validated scales and tests.
Studies should include documentation of the names of the scales and tests selected, reference(s), characteristics of the scale, and psychometric properties.

IR-5: Provide sufficient information in reports to allow for assessments of the study's internal and external validity.
Reporting guidelines for specific designs can be found at the EQUATOR Network website (www.equator-network.org). This website lists all reporting guidelines that have been developed using formal approaches, many of which have been adopted by journals, such as CONSORT (for randomized clinical trials), STARD (for diagnostic tests), STROBE (for observational studies), and SRQR and/or COREQ (studies using qualitative research). Researchers should register their studies with the appropriate registry (e.g., clinicaltrials.gov for clinical studies or observational outcomes studies) and provide complete and accurate responses to the information requested (e.g., enter the required and optional data elements for clinicaltrials.gov).

IR-6: Masking should be used when feasible.
Masking (also known as blinding) of research staff should be implemented, especially in situations for which study participant and investigator masking are not feasible. When masking is not feasible, the impact of lack of masking on the results should be discussed.

4: STANDARDS FOR PREVENTING AND HANDLING MISSING DATA

MD-1: Describe methods to prevent and monitor missing data.

Investigators should explicitly state potential reasons that study data may be missing. Missing data can occur from patient dropout, nonresponse, data collection problems, incomplete data sources, and/or administrative issues. As relevant, the protocol should include the anticipated amount of and reasons for missing data, plans to prevent missing data, and plans to follow up with study participants. The study protocol should contain a section that addresses steps taken in study design and conduct to monitor and limit the impact of missing data. This standard applies to all study designs for any type of research question.

MD-2: Use valid statistical methods to deal with missing data that properly account for statistical uncertainty due to missingness.

Valid statistical methods for handling missing data should be prespecified in study protocols. The analysis should explore reasons for missing data and assess the plausibility of the assumptions associated with the statistical methods. The potential impact of missing data on the results and limitations of the approaches used to handle the missing data should be discussed.

Estimates of treatment effects or measures of association should be based on statistical inference procedures that account for statistical uncertainty attributable to missing data. Methods used for imputing missing data should produce valid confidence intervals and permit unbiased inferences based on statistical hypothesis tests. Bayesian methods, multiple imputation, and various likelihood-based methods are valid statistical methods for dealing with missing data. Single imputation methods, such as last observation carried forward, baseline observation carried forward, and mean value imputation, are discouraged as the primary approach for handling missing data in the analysis. If single imputation-based methods are used, investigators must provide a compelling scientific rationale as to why the method is appropriate. This standard applies to all study designs for any type of research question.

MD-3: Record and report all reasons for dropout and missing data, and account for all patients in reports.

Whenever a participant drops out of a research study, the investigator should document the following: (1) the specific reason for dropout, in as much detail as possible; (2) who decided that the participant would drop out; and (3) whether the dropout involves participation in all or only some study activities. Investigators should attempt to continue to collect information on key outcomes for participants unless consent is withdrawn. All participants included in the study should be accounted for in study reports, regardless of whether they are included in the analyses. Any planned reasons for excluding participants from analyses should be described and justified. In addition, missing data due to other mechanisms (such as nonresponse and data entry/collection) should be documented and addressed in the analyses.

MD-4: Examine sensitivity of inferences to missing data methods and assumptions, and incorporate into interpretation.

Examining sensitivity to the assumptions about the missing data mechanism (i.e., sensitivity analysis) should be a mandatory component of the study protocol, analysis, and reporting. This standard applies to all study designs for any type of research question. Statistical summaries should be used to describe missing data in studies, including a comparison of baseline characteristics of units (e.g., patients, questions, or clinics) with and without missing data. These quantitative results should be incorporated into the interpretation of the study and reflected in the discussion section and, when possible, the abstract of any reports.

5: STANDARDS FOR HETEROGENEITY OF TREATMENT EFFECTS (HTE)

HT-1: State the goals of HTE analyses, including hypotheses and the supporting evidence base.
State the inferential goal of each HTE analysis, and explain how it is related to the topic of the research. Specify whether the HTE analysis is hypothesis driven (sometimes denoted as confirmatory), or hypothesis generating (sometimes denoted as exploratory). Hypothesis-driven HTE analyses should be prespecified, based on prior evidence (described clearly in the study protocol and study reports), and supported by a clear statement of the hypotheses the study will evaluate, including how subgroups will be defined (e.g., by multivariate score or stratification), outcome measures, and the direction of the expected treatment effects.

HT-2: For all HTE analyses, provide an analysis plan, including the use of appropriate statistical methods.
The study protocol should unambiguously prespecify planned HTE analyses. Appropriate methods include, but are not limited to, interaction tests, differences in treatment effect estimates with standard errors, or a variety of approaches to adjusting the estimated subgroup effect, such as Bayesian shrinkage estimates. Appropriate methods should be used to account for the consequences of multiple comparisons; these methods include, but are not limited to, p-value adjustment, false discovery rates, Bayesian shrinkage estimates, adjusted confidence intervals, or validation methods (internal or external).

HT-3: Report all prespecified HTE analyses and, at minimum, the number of post-hoc HTE analyses, including all subgroups and outcomes analyzed.
Both protocols and study reports must report the exact procedures used to assess HTE, including data mining or any automatic regression approaches. HTE analyses should clearly report the procedures by which subgroups were defined and the effective number of subgroups and outcomes examined. Within each subgroup level, studies should present the treatment effect estimates and measures of variability. Prespecified HTE analyses (hypothesis driven) should be clearly distinguished from post-hoc HTE analyses (hypothesis generating). Statistical power should be calculated and reported for prespecified (hypothesis-driven) analyses.

6: STANDARDS FOR DATA REGISTRIES

DR-1: Requirements for the design of registries
Registries established for conducting patient-centered outcomes research (PCOR) must have the following characteristics:

A. **Registry Purpose and Protocol.** The purpose of the registry should be clearly defined to guide the design of key registry features including, but not limited to, the target population, the research question(s) to be addressed, the data source used, the data elements collected, data sharing policies, and the stakeholders involved in the development and use of the registry. Participants and other key stakeholders should be engaged in registry design and protocol development. Registries should aim to be user oriented in design and function.

B. **Data Safety and Security.** Registry custodians should comply with institutional review board (IRB) human subjects protection requirements, the HIPAA Privacy Rule, and all other applicable local, state, and national laws. Registries should provide information describing the type of data collection (primary or secondary source data), data use agreements (DUAs), informed consent documents, data security protections, plans for maintaining data protection if the registry ends, and approaches to protecting privacy, including risk of and/or process for re-identification of participants, especially for medical or claims records.

C. **Data Elements and Quality.** Standardized data element definitions and/or data dictionaries should be used whenever possible. When creating a new registry, published literature should be reviewed to identify existing, widely used definitions of outcomes, exposures, and confounders before drafting new definitions.

When collecting primary data, conduct multistakeholder engagement with potential participants and data users to prioritize data collection needs. When participants support their face validity, use validated instruments or PRO measures when available. If secondary data sources (e.g., electronic medical records, claims data) are used, describe the original purpose of the secondary data and verify the accuracy and completeness of the data, as well as the approach to and validity of the linkages performed between the primary and secondary sources.

The specifics of the quality assurance plan will depend on the type of data (primary or secondary) collected by the registry. In general, the plan should address (1) structured training tools for data abstractors/curators; (2) the use of data quality checks for ranges and logical consistency for key exposure and outcome variables and covariates; and (3) data review and verification procedures, including source data verification plans (where feasible and appropriate), and validation statistics focused on data quality for the key exposure and outcome variables and key covariates. A risk-based approach to quality assurance, focused on variables of greatest importance, is advisable.

D. Confounding. Registries should identify important potential confounders pertinent to the purpose and scope of the research during the planning phase and collect reasonably sufficient data on these potential confounders to facilitate the use of appropriate statistical techniques during the analysis phase. When conducting analyses, refer to the PCORI Methodology Standards for Data Integrity and Rigorous Analyses and Standards for Causal Inference Methods.

E. Systematic Participant Recruitment and Enrollment. Develop a sampling plan of the target population and identify recruitment strategies for participants that minimize the impact of selection bias. Participants should be enrolled systematically, with similar procedures implemented at all participating sites and for each intervention of interest. Confirm adherence to agreed-upon enrollment practices.

F. Participant Follow-Up. The objective(s) of the registry should determine the type, extent, and length of participant follow-up.

Describe the frequency with which follow-up measures will be ascertained, consider linkage with other data sources (e.g., the National Death Index) to enhance long-term follow-up, and identify the date of last contact with the participant in existing registries, where appropriate. Ensure that the participants are followed in as unbiased a manner as possible, using similar procedures at all participating sites.

Monitor loss to follow-up to ensure best efforts are used to achieve follow-up time that is adequate to address the main objective. At the outset of the registry, develop a retention plan that documents when a participant will be considered lost to follow-up and which actions will be taken to minimize loss of pertinent data. Retention efforts should be developed with stakeholders to ensure the efforts are suitable for the target population and anticipated challenges are addressed appropriately.

DR-2: Documentation and reporting requirements of registry materials, characteristics, and bias

Clearly describe, document with full citations where appropriate, and make publicly available registry materials including, but not limited to, registry protocols, data-sharing policies, operational definitions of data elements, survey instruments used, and PROs captured. Modifications to any documents or data collection instruments should be clearly described and made available for registry users and participants. Characteristics of the participants in the registry should be described. Identify how the participants may differ from the target population to help assess potential selection biases. Document the loss to follow-up and describe the impact on the results, using sensitivity analyses (prespecified where possible) to quantify possible biases. Report the extent of bias clearly to stakeholders who may want to use the registry resource.

DR-3: Adapting established registries for PCOR

Previously established registries that intend to support new clinical research may not have been informed by all applicable methodology standards. When new research will use such registries, investigators should engage key stakeholders, including registry participants, to assess the feasibility of using the registry for new research and ensure the following:

- Informed consent documents are appropriately tailored to participant needs, characteristics, and conditions.
- Data elements are meaningful and useful to researchers and participants.
- Recruitment and retention strategies are feasible and effective.
- Registry policies are patient centered and the use of registry data is transparent to participants.
- Dissemination practices are appropriate and effective at reaching the communities from which the data are collected.
- Opportunities for bidirectional benefit exist between participants and researchers.
- Registry materials, described in **DR-2**, and informed consent forms are publicly available in accessible formats.

DR-4: Documentation requirements when using registry data

Researchers planning PCOR studies that rely on registries must ensure that these registries meet the requirements contained in Standards **DR-1** and **DR-2** and must document each required feature of each registry to be used (e.g., in an appendix to the funding application or study protocol). Deviations from the requirements with Standards DR-1 and DR-2 should be well documented and limitations of research related to the deviations from requirements should be addressed when reporting study findings.

7: STANDARDS FOR DATA NETWORKS AS RESEARCH-FACILITATING STRUCTURES

DN-1: Requirements for the design and features of data networks

Data networks established for conducting PCOR must have the following characteristics to facilitate valid, useable data and to ensure appropriate privacy, confidentiality, and intellectual property (IP) protections:

A. Data Integration Strategy. In order for equivalent data elements from different sources to be harmonized (treated as equivalent), processes should be created and documented that either (1) transform and standardize data elements prior to analysis or (2) make transformation logic (including code and process documentation) available that can be executed when data are extracted. The selected approach should be based on an understanding of the research domain of interest.

B. Risk Assessment Strategy. Data custodians should measure the risk of re-identification of data and apply algorithms to ensure that the desired level of confidentiality is achieved to meet the need of the particular PCOR application. Data custodians should ensure that data privacy/consents of the original data source cover the intended usage of the data through the data network. Privacy protections, including which data will be released and how breaches are addressed, should be specified in the data use agreement. The physical security of the data and data platforms should be considered and addressed as well.

C. Identity Management and Authentication of Individual Researchers. Develop reliable processes for verifying and authenticating the credentials of researchers who are granted access to a distributed research network.

D. IP Policies. A research network should develop policies for the handling and dissemination of IP; networks should also have an ongoing process for reviewing and refreshing those policies. IP can include data, research databases, papers, reports, patents, and/or products resulting from research using the network. Guidelines should balance (1) minimizing impediments to innovation in research processes and (2) making the results of research widely accessible, particularly to the people who need them the most.

E. Standardized Terminology Encoding of Data Content. The data content should be represented with a clearly specified standardized terminology system to ensure that their meaning is unambiguously and consistently understood by parties using the data.

F. Metadata Annotation of Data Content. Semantic and administrative aspects of data contents should be annotated with a set of metadata items. Metadata annotation helps to correctly identify the intended meaning of a data element and facilitates an automated compatibility check among data elements.

G. Common Data Model. Individual data items should be organized into a standard structure that establishes common definitions and shows close or distant associations among variables. A common data model specifies necessary data items that need to be collected and shared across participating institutes, clearly represents the associations and relationships among data elements, and promotes correct interpretation of the data content.

DN-2: Selection and use of data networks

Researchers planning PCOR studies that rely on data networks must ensure that these networks meet the requirements contained in **DN-1**, and they must document the current maintenance status of the data network (e.g., currency of the data, level of data curation). Because different studies are expected to have different dependencies on various components of the data network, researchers should assess the appropriateness of the data in the network for a specific research study through the following activities:

A. Data content and conformance. Document what is actually needed for the research question and compare that to the sources in the network. Identify which data are best represented by the network's data sources and how they are included in the study. Ensure that the representations and values of the data to be used from the network are sufficient for addressing the research question.

B. Data quality. Assess the data quality for the data sources that will be used. It is especially important to assess data completeness and plausibility. Where data are incomplete, identify and assess potential biases for completeness and consider alternate sources. Assess plausibility by reviewing data value distributions and comparing additional data sources that would have expected concordance with the selected sources. Determine whether the data sources are of sufficient quality to be included in the analysis.

C. Sensitivity analyses. After the initial analysis is completed, perform sensitivity analyses on the data sources to test whether possible variations in data characteristics would affect the conclusions of the analysis. Specifically, measure the sensitivity of the conclusions to the following:

- o Completeness and correctness of the data in the data network
- o Availability of data sources that are most likely at risk of exclusion
- o Temporal dependence of the data
- o Operational definitions and decisions made to implement analysis

The results of these assessments should be documented and included with any findings from research studies using the data networks.

8: STANDARDS FOR CAUSAL INFERENCE METHODS

CI-1: Specify the causal model underlying the research question *(cross-cutting standard, applies to all PCOR/CER studies)*.

Researchers should describe the causal model relevant to the research question, which should be informed by the PICOTS framework: populations, interventions, comparators, outcomes, timing, and settings. The causal model represents the key variables; the known or hypothesized relationships among them, including the potential mechanisms of effect; and the conditions under which the hypotheses are to be tested. Researchers should use the causal model to determine whether and how the study can handle bias and confounding and the extent to which valid estimates of the effects of an intervention can be generated given the particular hypothesis, study design, analytical methods, and data source(s).

CI-2: Define and appropriately characterize the analysis population used to generate effect estimates.

Researchers should specify the eligibility criteria for inclusion in the study population and analysis. Decisions about which patients are included in an analysis should be based on information available at each patient's time of study entry in prospective studies or on information from a defined time period prior to the exposure in retrospective studies. For

time-varying treatment or exposure regimes, specific time points should be clearly specified; relevant variables measured at baseline and up to, but not beyond, those time points should be used as population descriptors. When conducting analyses that in some way exclude patients from the original study population, researchers should describe the final analysis population that gave rise to the effect estimate(s), address selection bias that may be introduced by excluding patients, and assess the potential impact on the validity of the results.

CI-3: Define with the appropriate precision the timing of the outcome assessment relative to the initiation and duration of exposure.
To reduce potential sources of bias arising from inappropriate study design choices (e.g., immortal time bias), researchers must precisely define, to the extent possible, the timing of the outcome assessment relative to the initiation and duration of the exposure.

CI-4: Measure potential confounders before start of exposure and report data on potential confounders with study results.
In general, variables used in confounding adjustment (either in the design or analysis) should be ascertained and measured prior to the first exposure to the interventions (or intervention) under study. If confounders are time varying, specific time points for the analysis of the exposure effect should be clearly specified and the confounder history up to, and not beyond, those time points should be used in that analysis.

CI-5: Report the assumptions underlying the construction of propensity scores and the comparability of the resulting groups in terms of the balance of covariates and overlap.
When conducting analyses that use propensity scores to adjust for measured confounding, researchers should consider and report how propensity scores will be created (high dimensional propensity score versus a priori clinical variables) and which balancing method will be used (e.g., matching, weighting, or stratification). Researchers should assess and report the overlap and balance achieved across compared groups with respect to potential confounding variables.

CI-6: Assess the validity of the instrumental variable (i.e., how the assumptions are met) and report the balance of covariates in the groups created by the instrumental variable.
When an instrumental variable (IV) approach is used (most often to address unmeasured confounding), empirical evidence should be presented that describes how the variable chosen as an IV satisfies the three key properties of a valid instrument: (1) the IV influences the choice of intervention or is associated with a particular intervention because both have a common cause; (2) the IV is unrelated to patient characteristics that are associated with the outcome; and (3) the IV is not otherwise related to the outcome under study (i.e., it does not have a direct effect on the outcome apart from its effect through exposure).

9: STANDARDS FOR ADAPTIVE AND BAYESIAN TRIAL DESIGNS

AT-1: Specify planned adaptations, decisional thresholds, and statistical properties of those adaptations.
The adaptive clinical trial design must be prospectively planned and the design must be clearly documented in the study protocol before trial enrollment begins, including at a minimum the following:

- All potential adaptations, including timing
- Interim trial findings that will be used in determining each adaptation
- Statistical models and decisional thresholds to be used
- Planned analyses of the trial endpoint(s)

The description of the design should be sufficiently detailed that it could be implemented based on the description of

procedures. This specification should include a statistical analysis plan in which all necessary detail is provided regarding planned interim and final analyses.

Additionally, the statistical properties of adaptive clinical trial designs should be thoroughly investigated over the relevant range of important parameters or clinical scenarios (e.g., treatment effects, accrual rates, delays in the availability of outcome data, dropout rates, missing data, drift in participant characteristics over time, subgroup-treatment interactions, or violations of distributional assumptions). Statistical properties to be evaluated should include Type I error, power, and sample size distributions, as well as the precision and bias in the estimation of treatment effects.

AT-2: Specify the structure and analysis plan for Bayesian adaptive randomized clinical trial designs.

If a Bayesian adaptive design is proposed, the Bayesian structure and analysis plan for the trial must be clearly and completely specified. This should include any statistical models used either during the conduct of the trial or for the final analysis, prior probability distributions and their basis, utility functions associated with the trial's goals, and assumptions regarding exchangeability (of participants, of trials, and of other levels). Specific details should be provided as to how the prior distribution was determined and if an informative or noninformative prior was chosen. When an informative prior is used, the source of the information should be described. If the prior used during the design phase is different from the one used in the final analysis, then the rationale for this approach should be indicated. Computational issues should be addressed, including describing the choice of software, the creation and testing of custom software, and software validation. Software used for Bayesian calculations during trial design, trial execution, and final analysis must be functionally equivalent. When feasible, software or other computing packages should be made available to relevant stakeholders for evaluation and validation.

AT-3: Ensure that clinical trial infrastructure is adequate to support planned adaptation(s) and independent interim analyses.

The clinical trial infrastructure, including centralized randomization, data collection related to the assessment and recording of key outcomes, data transmission procedures, and processes for implementing the adaptation (e.g., centralized, web-based randomization), must be able to support the planned trial. In simple adaptive trials, qualitative verification of the capabilities of the proposed trial infrastructure may be adequate. Trials with more complicated requirements, such as frequent interim analyses, require thorough testing prior to trial initiation. Such testing should involve the trial's data collection and data management procedures, the implementation of the adaptive algorithm, and methods for implementing the resulting adaptation(s). The impact on the trial's operating characteristics of delays in collecting and analyzing available outcome data should be assessed. The study plan should clarify who will perform the analyses to inform adaptation while the study is ongoing and who will have access to the results. The interim analyses should be performed and reviewed by an analytical group that is independent from the investigators who are conducting the trial. Trial investigators should remain blinded to changes in treatment allocation rates as this information provides data regarding treatment success.

AT-4: When reporting adaptive randomized clinical trials, use the CONSORT statement, with modifications.

The following sections of the 2010 CONSORT statement can be used to report key dimensions of adaptation:

- Adaptation of randomization probabilities (sections 8b and 13a)
- Dropping or adding study arms (sections 7b and 13a)
- Interim stopping for futility and superiority or adverse outcomes (sections 7b and 14b)
- Sample size re-estimation (sections 7a and 7b)
- Transitioning of stages (e.g., seamless Phase II/III designs) (sections 3a, 7a, 7b, and 16)
- Modification of inclusion and exclusion criteria (sections 4a and 13a)

CONSORT sections 16, 20, and 21 provide additional guidance on reporting aspects of an adaptive trial.

All possible adaptations included in the prospective design, even if they did not occur, should be included in the study reports.

10: STANDARDS FOR STUDIES OF MEDICAL TESTS

(formerly Standards for Studies of Diagnostic Tests)

MT-1: Specify the clinical context and key elements of the medical test.

Evaluation of tests used to inform medical decision making (e.g., diagnostic, prognostic, or predictive tests) should specify each of the following items and provide justification for the particular choices: (1) the intended use of the test and the corresponding clinical context, including referral for additional testing, referral for additional treatments, and modification of current treatment and target populations; (2) the choice of comparator (e.g., another test or no test) and goal of the comparison; (3) the technical specifications of the test(s) as implemented in the study; (4) the approach to test interpretation; (5) the sources and process for obtaining reference standard information, when applicable; (6) the procedures for obtaining follow-up information and determining patient outcomes, when applicable; and (7) the clinical pathways involving the test and the anticipated implications of test use on downstream processes of care and patient outcomes. These items ought to be specified for all types of tests used for medical decision making and for all designs, including observational designs (e.g., those using medical records or registries). If these items are not available directly, validated approaches to approximating these study elements from available data should be used.

MT-2: Assess the effect of factors known to affect performance and outcomes.

Studies of tests used to inform medical decision making should include an assessment of the effect of important factors known to affect test performance and outcomes, including, but not limited to, the threshold for declaring a "positive" test result, the technical characteristics of the test, test materials (e.g., collection, preparation, and handling of samples), operator dependence (e.g., lab quality, interpretation requirements), and the setting of care.

MT-3: Focus studies of medical tests on patient-centered outcomes, using rigorous study designs with a preference for randomized controlled trials.

A prospective randomized design should be used when possible to assess the diagnostic, prognostic, predictive, and/or therapeutic outcomes of testing. If a nonrandomized design is proposed, a rationale for using an observational study (or modeling and simulation) should be provided, and efforts to minimize confounding documented.

11: STANDARDS FOR SYSTEMATIC REVIEWS

SR-1: Adhere to National Academy of Medicine (NAM) standards for systematic reviews of comparative clinical effectiveness research, as appropriate.

Systematic reviews, which critique and synthesize the existing literature, can also identify evidence gaps and inform decisions of how to address these gaps. Existing standards for systematic reviews developed by credible authorities, such as the Cochrane Collaboration and the Agency for Healthcare Research and Quality, vary somewhat in their recommended approaches. The PCORI Methodology Committee endorses the standards issued by the NAM in 2011 but recognizes both the importance of conducting systematic reviews consistent with updates to best methodological practices and that there can be flexibility in the application of some standards without compromising the validity of the review, including the following:

- Searches for studies reported in languages other than English are not routinely recommended but may be appropriate to some topics.
- Dual screening and data abstraction are desirable, but fact-checking may be sufficient. Quality control procedures are more important than dual review per se.
- Independent librarian peer review of the search strategy is not required; internal review by experienced researchers is sufficient.

Researchers should describe and justify any departures from the 2011 NAM standards (e.g., why a particular requirement does not apply to the systematic review).

12: STANDARDS ON RESEARCH DESIGNS USING CLUSTERS

RC-1: Specify whether the study objectives, the interventions, and the primary outcomes pertain to the cluster level or the individual level.
Describe (1) the target population of clusters and individuals to which the study findings will be generalizable and (2) the clusters to be randomized and the subjects to be enrolled in the trial.

RC-2: Justify the choice of cluster randomization.
Describe the benefits and disadvantages of cluster randomization versus individual-level randomization for the proposed research. Cluster randomization should be substantiated by a sound theoretical and conceptual framework that describes the hypothesized causal pathway (see **CI-1**). Cluster randomization generally is applicable in the following instances:

- An intervention is delivered at the cluster level.
- An intervention changes the physical or social environment.
- An intervention involves group processes.
- An intervention cannot be delivered without a serious risk of contamination.

Logistical considerations can also justify cluster randomization, for example to reduce costs or to improve participation, adherence, or administrative feasibility.

RC-3: Power and sample size estimates must use appropriate methods to account for the dependence of observations within clusters and the degrees of freedom available at the cluster level.
The methods used to reflect dependence should be clearly described. Sources should be provided for the methods and for the data used to estimate the degree of dependence. Sensitivity analyses incorporating different degrees of dependence must be reported. For simpler designs, the dependence in the data can be reflected in the intraclass correlation. Dependence can also be reflected in variance components. Other factors that affect the power calculation and should be described include the design of the study, the magnitude of the hypothesized intervention effect, the prespecified primary analysis, and the desired Type I error rate.

RC-4: Data analyses must account for the dependence of observations within clusters regardless of its magnitude.
Data analyses must also reflect the degrees of freedom available at the cluster level. Investigators must propose appropriate methods for data analyses with citations and sufficient detail to reproduce the analyses.

RC-5: Stratified randomization should be used when feasible.
Because cluster randomization trials often involve a limited number of groups or clusters, stratified randomization should be considered and is recommended when feasible. If not feasible, justification should be provided for the use of other methods. The recommended stratification factors are those that are expected to be strongly correlated with the outcome or with the delivery of the intervention, such as baseline value of the outcome variable, cluster size, and geographic area.

Only a limited number of confounders can be addressed through stratification. Other variables, particularly those that characterize the context, should be measured and assessed to document their potential influence on the outcome and understanding of heterogeneity of results.