Flatiron Health - Academic Research Proposal [Template]

*Updated April 2021*

**INSTRUCTIONS:** This template is to be completed by academic investigators interested in completing a research project using a nationwide electronic health record-derived dataset from Flatiron Health. The goal of this form is to provide sufficient detail to inform early exploration of project feasibility by clearly outlining the target cohorts, data variables, and high-level analytic plan.

|  |  |
| --- | --- |
| **Category** | **Description** |
| **Working project title** |  |
| **Principal Investigator (PI) name, affiliation, and contact information***Fellows and trainees may serve as co-PI, below, but are not eligible for the role of PI. The PI faculty member must be included in the development of the research proposal and confirm their support for the proposal prior to submission.* |  |
| **Co-Investigator (co-PI) and collaborator names, affiliations, and contact information***Indicate if a co-PI is the contact submitting the application. Include any industry collaborators and collaborators at other academic institutions, if applicable.**If you or any members of your research team need access to Flatiron Explore, please complete* [*this form*](https://docs.google.com/forms/d/1d93vWs_YMXEMkC1svaTJXeEAVOnAY-4X0eyTu_8DXt4/edit?usp=drive_web) *and we will fulfill your request within 5-10 business days.* |  |
| **Data scientist coding the analysis, name, affiliation, and contact information***Due to the complexity of Flatiron data, a data scientist is an essential part of the investigative team.**This role should be filled by a biostatistician, epidemiologist, economist, software engineer, or other person with technical training in data wrangling and reproducible coding of analyses.* | *Please also describe relevant training and experience working with large data sets (e.g. Flatiron, SEER-Medicare, NCDB).* |
| **Disease(s) of interest***Resources on Flatiron Explore:** [*Explore data*](https://academics.flatiron.com/explore/explore_data) *to learn about each disease*
 |  |
| **Background and objectives** | *Summarize the main findings of a critical review of the literature pertinent to the study objectives. Describe what is currently known and highlight gaps in knowledge the study will address.****Topics (as applicable):***Background:* *General background about the disease and characteristics of the specific patient population relevant to the study*

Knowledge gap:* *Highlight gaps in knowledge that this study will address.*

Treatments or exposures:* *Name and description of any treatments/therapies of interest to the study, as applicable;*
* *Provide information regarding safety and/or effectiveness of the treatments/therapies of interest, as applicable;*

Objectives:* *Provide overall summary of study objectives*

References:* *References to the literature and data that are relevant to the study*
 |
| **Research question**(1-2 sentences) | *State the overarching research question the study will address with a clear tie to the gaps in knowledge highlighted in the Background section.* |
| **Specific aims and hypothesis**Primary, secondary, exploratory*Describe concisely each of the specific aims and corresponding hypothesis of the analysis.* | *The aims should be specific and convey the study population, key variables, and potential statistical approach intended for that aspect of the analysis.**Hypotheses are required for hypothesis testing aims, but may not always be appropriate, especially for exploratory analyses.****Frequently reported (if applicable):***Primary study aim(s): Hypothesis:Secondary study aim(s): Hypothesis:Exploratory aim(s): Hypothesis: |
| **Cohort inclusion and exclusion criteria***Resources on Flatiron Explore:** [*Analytic guidance for academic investigators*](https://flatironacademics.zendesk.com/hc/en-us/articles/360015450153-Analytic-Considerations-for-Academic-Investigators)
* [*90 day gap rule methodology tutorial*](https://flatironacademics.zendesk.com/hc/en-us/articles/360022330814)
* [*Selection of your study population*](https://flatironacademics.zendesk.com/hc/en-us/articles/360021912254-Selection-of-the-study-population)
 | *Explicitly describe the inclusion and exclusion criteria, using variables available in the Flatiron Health national datasets, which can be viewed in the disease-specific data dictionaries in* [*Flatiron Explore*](https://academics.flatiron.com/explore/explore_data)*. Please note that if any I/E criteria are specified using data elements that are not found in these data dictionaries, we will be unable to move forward with approving the feasibility of your research proposal.**If additional sub-cohorts are being constructed, the additional criteria dictating creation of these sub-cohorts should also be described. Include a cohort selection diagram, if appropriate. This could be a shell of a cohort selection diagram (without numbers) explaining what steps will be taken during the project.****Frequently reported criteria (if applicable):***Inclusion criteria* *Study population (e.g., metastatic breast cancer patients)*
* *Age range (e.g., all adult patients ages 18+ years)*
* *Sex (e.g., restricted to female breast cancers)*
* *Cohort selection*
	+ *Treatment assignment (e.g., first-line therapy with CDK4/6 inhibitor plus endocrine therapy)*
	+ *Clinical characteristics (e.g., documented ECOG performance status)*
	+ *Disease-specific characteristics (e.g., de novo advanced cancers)*
	+ *Biomarker status (e.g., documented FISH testing)*
	+ *Enhanced cohort characteristics (e.g., M-spike testing available to measure progression-free survival)*

Exclusion criteria* *Missing potential treatment information (e.g.,* [*90 day gap rule*](https://flatironacademics.zendesk.com/hc/en-us/articles/360022330814) *or other relevant rules)*
* *Minimum follow up (e.g. at least 6 months since metastatic diagnosis date)*
* *Minimum treatment assignment (e.g., at least 1L)*

Observation period:* *Dates of observation (e.g., index date between March 31, 2017 and data cutoff)*
* *Index time (e.g., start date of first-line therapy)*
 |
| **Data elements required**Variables requested for this analysis (including key clinical descriptors, endpoints) *Please use variable and/or table names in the Flatiron Health data dictionaries available for download in* [*Flatiron Explore*](https://academics.flatiron.com/research/getting_started) | *Include a list of all variables available in the Flatiron Health national datasets to be included in the analysis, including any covariates or outcomes / endpoints of interest. Describe any important and relevant details, such as how they will be defined, measured, and operationalized (including things like handling of special cases and variable transformation).**Please note that if any variables are requested that are not found in the disease-specific data dictionaries in* [*Flatiron Explore*](https://academics.flatiron.com/explore/explore_data)*, we will be unable to move forward with approving the feasibility of your research proposal.*Enhanced Table (e.g. enhanced\_metastaticbreast):* *Patientid*
* *Diagnosisdate*
* *Metdiagnosisdate*
* *GroupStage*
* *...*

Demographics Table:* *Birthyear*
* *Gender*
* *Race*
* *Ethnicity*
* *Practicetype*
* *...*

ECOG Table:* *ECOG value*
* *ECOG date*
* *...*

LineOfTherapy Table:* *Linenumber*
* *Linename*
* *Startdate*
* *Enddate*
* *...*

Drugepisode Table:* *Linename*
* *Linenumber*
* *Ismaintainancetherapy*
* *Episodedate*
* *Drugname*
* *Detaileddrugcategory*
* *…*

Biomarkers Table:* *Biomarkername*
* *Specimencollecteddate*
* *Specimenreceiveddate*
* *Resultdate*
* *Biomarkerstatus*
* *Testype*
* *...*

Visit Table*…* |
| **Statistical analysis plan** *Describe the approach and methods to account for bias in non-randomized study. See the Flatiron Health* [*analytic guidance for academic investigators*](https://flatironacademics.zendesk.com/hc/en-us/articles/360015450153-Analytic-Considerations-for-Academic-Investigators)*. Please include power calculations if appropriate.**Recommended* [*reference*](https://pubmed.ncbi.nlm.nih.gov/23940288/) *and guidance for handling threats to validity in longitudinal follow up studies in cancer patients.* | ***Please include a detailed statistical analysis plan for each aim.****Describe any important and relevant details, such as how key variables will be defined, measured, and operationalized (including things like handling of special cases and variable transformation). Indicate the statistical software intended to be used for the analysis. Any sensitivity analyses should be described, as should any planned subgroup analyses. Investigators are encouraged to review Flatiron’s analytic guidance (link in left column).****Frequently reported elements (examples):*****Descriptive Analyses***The statistical analysis section should include a description of the descriptive statistics to be generated.** *The comparison groups should be explicitly categorized.*
	+ *Between group comparisons of continuous variables (e.g., multi-way ANOVA, Wilcoxon rank-sum test)*
	+ *Between group comparisons of categorical variables (e.g., chi-square and/or Fisher’s exact test)*
* *The estimated trends over time should be explicitly defined.*
	+ *Proportions should be used (e.g. proportion of patients treated with 1L immunotherapy among patients treated with any 1L), instead of absolute counts due to Flatiron cohort sampling methodology.*
	+ *Inference should be drawn with careful consideration of potential missingness and bias.*

**Hypothesis Testing Analyses***The statistical methodology to be applied to test each hypothesis should be specified, along with the rationale for using the methodology. Where the proposed statistical methodology includes modeling the data, methods to ensure that the model assumptions are fulfilled should be described.* Multivariable regression modeling* *Outcome (e.g., type of first-line treatment initiated)*
* *Main exposure (e.g. race/ethnicity)*
* *Covariates (e.g. age at metastatic diagnosis [continuous], ECOG PS at metastatic diagnosis [0, 1, 2, 3, 4, missing], …)*
* *Type of model (e.g., negative binomial regression)*
	+ *Estimates (e.g., RR and 95% CI)*
* *Evaluation of model fit and covariate selection if applicable*
* *Interaction analysis*

 Time-to-event (survival) analyses:* *Index date (e.g. metastatic diagnosis date)*
* *Event (e.g., overall survival)*
* *Censoring (e.g., last clinic date)*
* *Main exposure or comparison group (e.g. 1L immunotherapy vs chemotherapy, where patients received 1L chemotherapy as reference group)*
* *Covariates (e.g. age at metastatic diagnosis [continuous], ECOG PS at metastatic diagnosis [0, 1, 2, 3, 4, missing], …)*
* *Evaluation of model assumptions (e.g. proportional hazards assumption for Cox PH model)*
	+ *Alternative approach if the assumption is violated*

Propensity score approaches utilizing matching and/or inverse probability of treatment weighting:* *Regression model (e.g., logit) for estimating propensity scores*
* *Variable selection for propensity score model*
* *Possible threats to validity and assumptions (i.e., consistency, exchangeability, positivity, and no misspecification)*
* *Balance diagnostics (e.g., comparison of means and proportions of baseline variables)*
* *IPTW diagnostics (e.g., comparison of means prevalences in the weighted sample)*
* *Matching diagnostics (e.g., plotting density functions for distribution balance of key variables)*

Power calculations and sample size:* *Anticipated cohort size(s) [suggest use Flatiron Explore -> Explore Data -> Access Feasibility for cohort size estimates if applicable]*
* *Minimum detectable differences/effect estimates*
* *Power (defined a priori)*
 |
| **Describe potential limitations and biases and risk mitigation strategies** | *Include information regarding any potential limitations of the study design, data sources, and analytic methods, including issues relating to confounding, bias, generalizability, and random error. Investigators are encouraged to review Flatiron’s analytic guidance (link above).****Frequently reported elements (examples):***Handling of missing data:* *Exclusions (e.g., no documented FISH testing for biomarker-defined cohort)*
* *Approaches to missing covariate data (e.g., multiple imputation by chained equations for >20% missing)*

Approaches to bias and confounding:* *Confounders measured a priori*
* *Potential selection biases and mitigation strategy (e.g., accounting for left truncation using uniform start date)*
* *Potential information biases and mitigation strategy (e.g., accounting for immortal time bias using time-dependent exposure)*

Multiple testing:* *Adjustment for multiple testing (e.g. Bonferroni correction, false-discovery rate correction)*
 |
| **Expected significance and impact of study results** | *Briefly describe the expected contribution of this study to the general knowledge about this specific treatment/therapy or patient population.* |
| **Project timeline and intended use of study results**(e.g., grant application\*, conference abstract, manuscript publication)*Include any relevant interim milestones, target dates, and submission deadlines**\*Note: For grant proposal requests, ,ore information can be found in* [*Flatiron Explore*](https://flatironacademics.zendesk.com/hc/en-us/articles/360026669013) |  |
| **Plan for IRB review and approval***Investigators must confirm local IRB requirements and submit an IRB approval letter or waiver to Flatiron Health prior to data delivery. Flatiron can provide documentation of central IRB approval of research processes upon request, but this does not replace local review requirements.* |  |
| **Anticipated funding source**(if applicable) | *If applicable, add existing funding and/or potential funding application plans (e.g. Institutional K grant, NCI R-01 application for spring or fall standard due date).* |
| **Additional information**(if applicable) | *If applicable, provide additional materials, including but not limited to:** *Table shells/mock plots*
* *Details of statistical analysis*
* *Cohort selection diagram*
 |

**Notes about requests for additional abstraction:**

Please note each disease cohort in the standard Flatiron Health data model has its own set of variables, which can be viewed in the disease-specific data dictionaries in Flatiron Explore. In order to accommodate research projects that require data outside of the standard Flatiron Health data models, the Flatiron Health team has created a list of standard variables (see below) that are available for abstraction contingent upon researcher funding for the abstraction. If you are interested in requesting additional abstraction please contact Flatiron Health at academic-partnerships@fltairon.com.

* Detailed orals
* Comorbidities (e.g., CCI)
* Reason for final discontinuation
* Additional malignancies
* Sites of metastases\*

\* Note: some of our disease models include sites of metastases in our standard data offering. Please refer to the data dictionaries or [this article](https://flatironacademics.zendesk.com/hc/en-us/articles/360010106734) within Flatiron Explore for more information.

**Notes about project overlap:**

As with other real-world data sources used for research, there is a possibility this proposal overlaps with other ongoing research projects. Flatiron seeks to minimize overlap whenever possible by providing high-level visibility into approved research projects submitted by our academic partners within Flatiron Explore [research projects](https://academics.flatiron.com/explore/research_projects/all).