Projecting enrollment in a pragmatic trial that utilizes real-world data

Devin Incerti
Genentech

Acknowledgements

This work reflects the efforts of a large number of collaborators from Flatiron Health, Foundation Medicine (FMI), Genentech, and the clinical sites used in the study. Brian Segal, Arjun Sondhi, and Guneet Wallia, in particular, were instrumental in helping develop and implement the methodology that will be presented today.

What are pragmatic trials?

Aim to estimate the effects of treatments in routine clinical practice

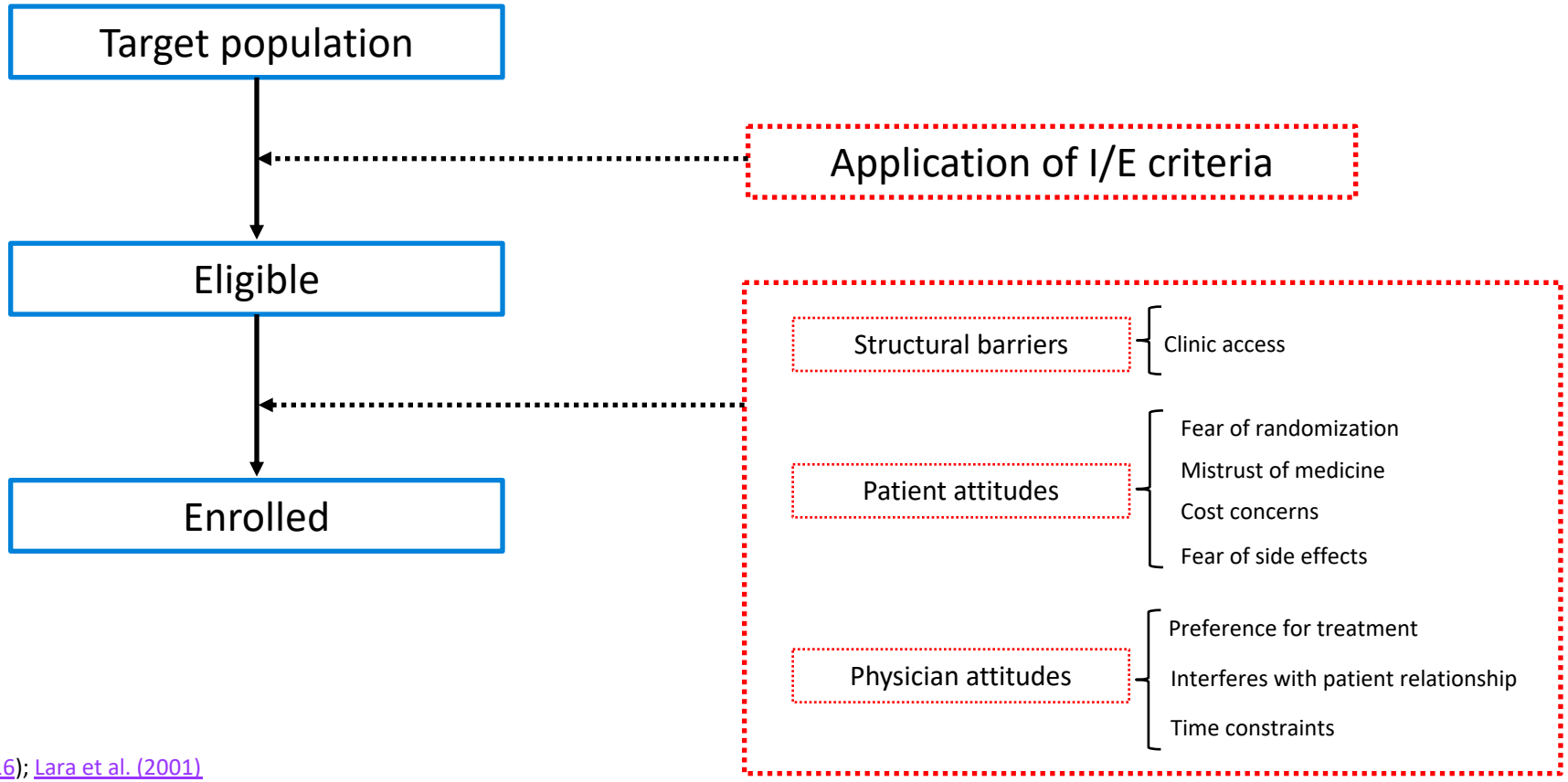
	Clinical trial	Pragmatic trial
Eligibility	Stringent inclusion / exclusion criteria	Simplified inclusion / exclusion criteria
Setting	Academic centers	Community sites
Follow-up	Frequent visits & assessment; shorter-term	Part of routine practice; longer-term; can leverage RWD infrastructure (e.g., EHR)
Adherence	Measures to increase adherence (e.g., additional monitoring)	Patient treated according to current usual care
Intervention	Double blinded placebo controlled	Non-blinded SoC control

Why use pragmatic trials?

- Better external validity
- Larger sample size and longer follow-up
- Increased efficiency and feasibility; reduced R&D costs
- Address the needs of multiple stakeholders (e.g., regulators, clinicians, patients, payers)
- Can incorporate randomization to estimate relative treatment effects

Efficiency and feasibility

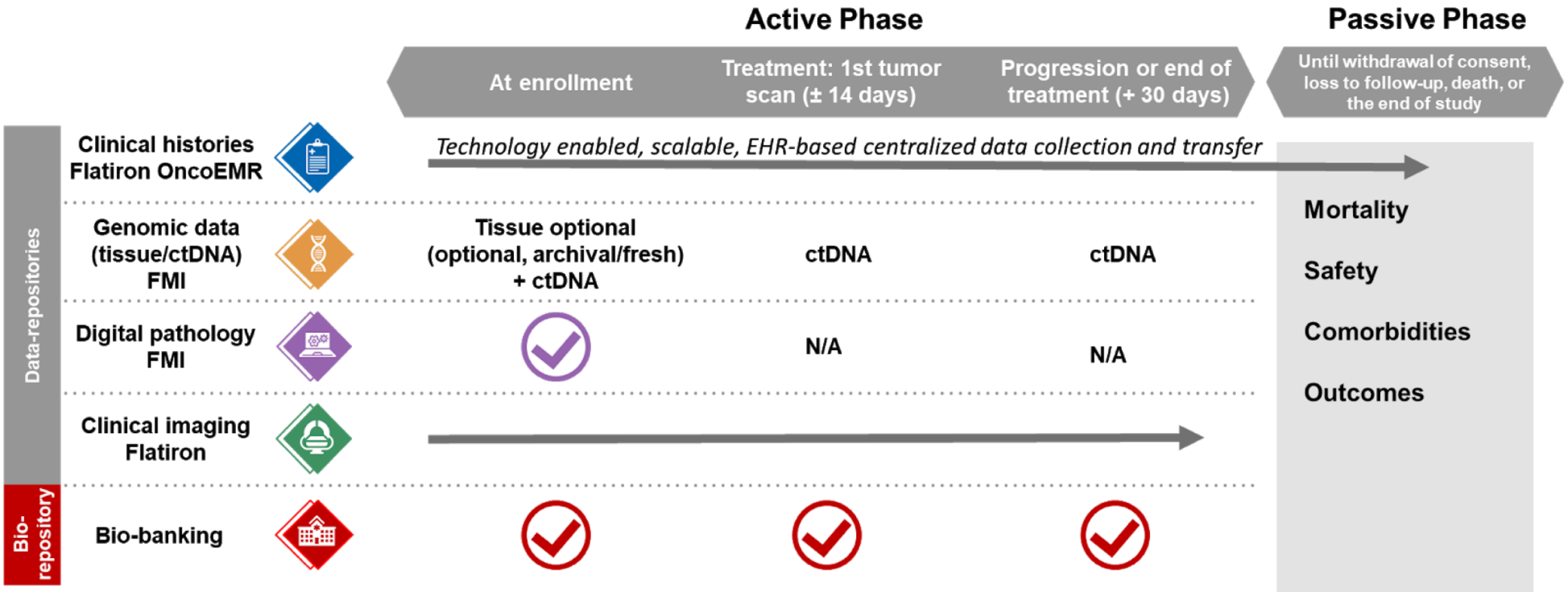
Estimates suggest that <5% of cancer patients participate in clinical trials[1]



[1]Unger et al. (2016); Lara et al. (2001)

The Flatiron/FMI/Genentech prospective clinico-genomic study (PCG)

- *Actively* collect circulating tumor DNA (ctDNA) at 3 defined time points and then *passively* collect data through existing EHR systems
- 1,000 patients with either mNSCLC or ES-SCLC
- *No randomization*



mNSCLC = metastatic non-small cell lung cancer; ES-SCLC = extensive stage small cell lung cancer

PCG has a less stringent inclusion/exclusion criteria than a typical clinical trial

(Example comparison with IMpower150)

PCG

Inclusion criteria

- Diagnosis of mNSCLC
- Planned initiation of SOC systemic anti-cancer treatment
- Age ≥ 18
- Ability to comply with study protocol

Exclusion criteria

- Receiving investigational medicinal product(s) as part of an interventional trial

IMpower150[1]

Inclusion criteria

- Stage IV non-squamous NSCLC
- Age ≥ 18
- Measurable disease as defined by RESIST
- Known PDL-1 status
- ECOG of 0 or 1
- Adequate hematologic and end organ function
- No prior treatment for stage IV non-squamous NSCLC

Exclusion criteria

Comorbidities

- Active or untreated central nervous system metastases
- Malignancies other than NSCLC within 5 years prior to randomization
- Pregnant or lactating women
- History of autoimmune disease
- HIV
- Active hepatitis B or C
- History of a number of lung diseases (e.g., pulmonary fibrosis, pneumonia, pneumonitis)
- Severe infection within 4 weeks prior to randomization
- Significant cardiovascular disease
- Illness that impacts ability to follow/comply study procedures

Prior treatment

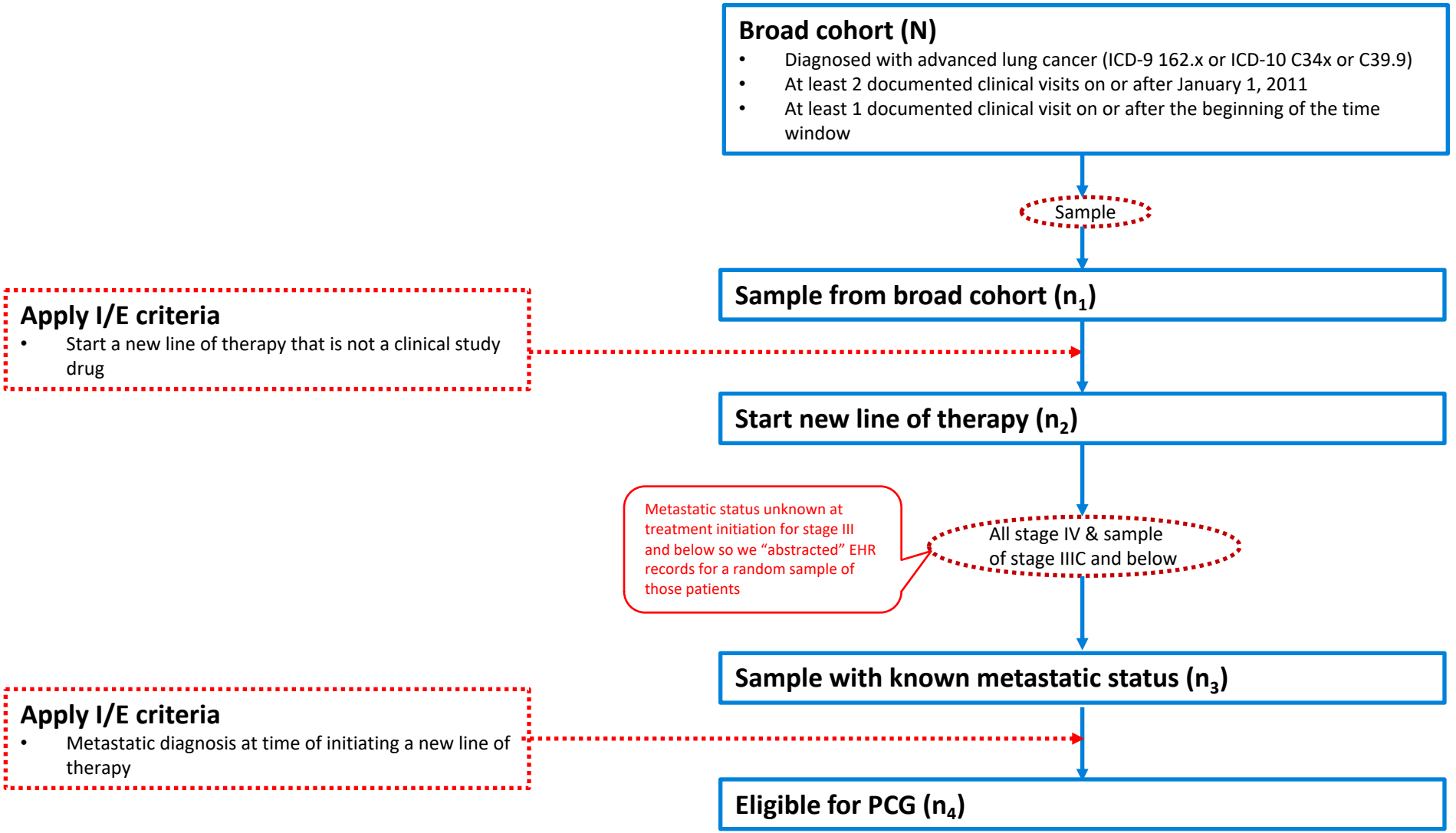
- Cluster of differentiation 137 agonists or immune checkpoint blockade therapies, anti-programmed death-1, and anti-PD-L1 therapeutic antibodies

[1]Randomized open-label study comparing atezolizumab with chemotherapy to chemotherapy alone

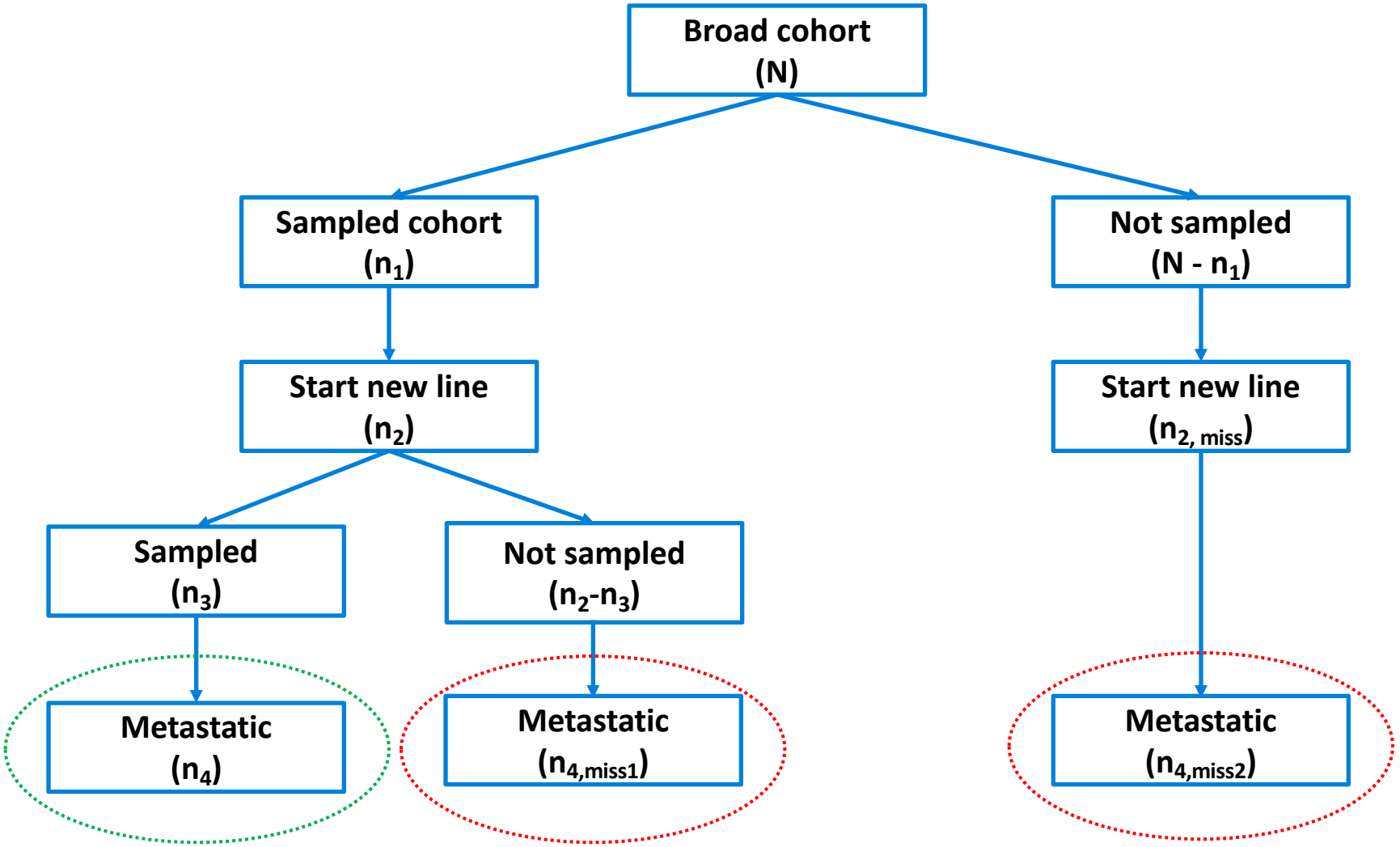
PCG primary analysis

Objectives	Assess feasibility, develop methods to aid planning & protocol extensions, and inform future pragmatic studies
Primary estimand	% of eligible patients that enrolled

Estimation of the number of eligible patients is surprisingly difficult because of a complex cohort selection process *(Illustration for mNSCLC)*



The cohort selection process implies the sampling scheme for a single site



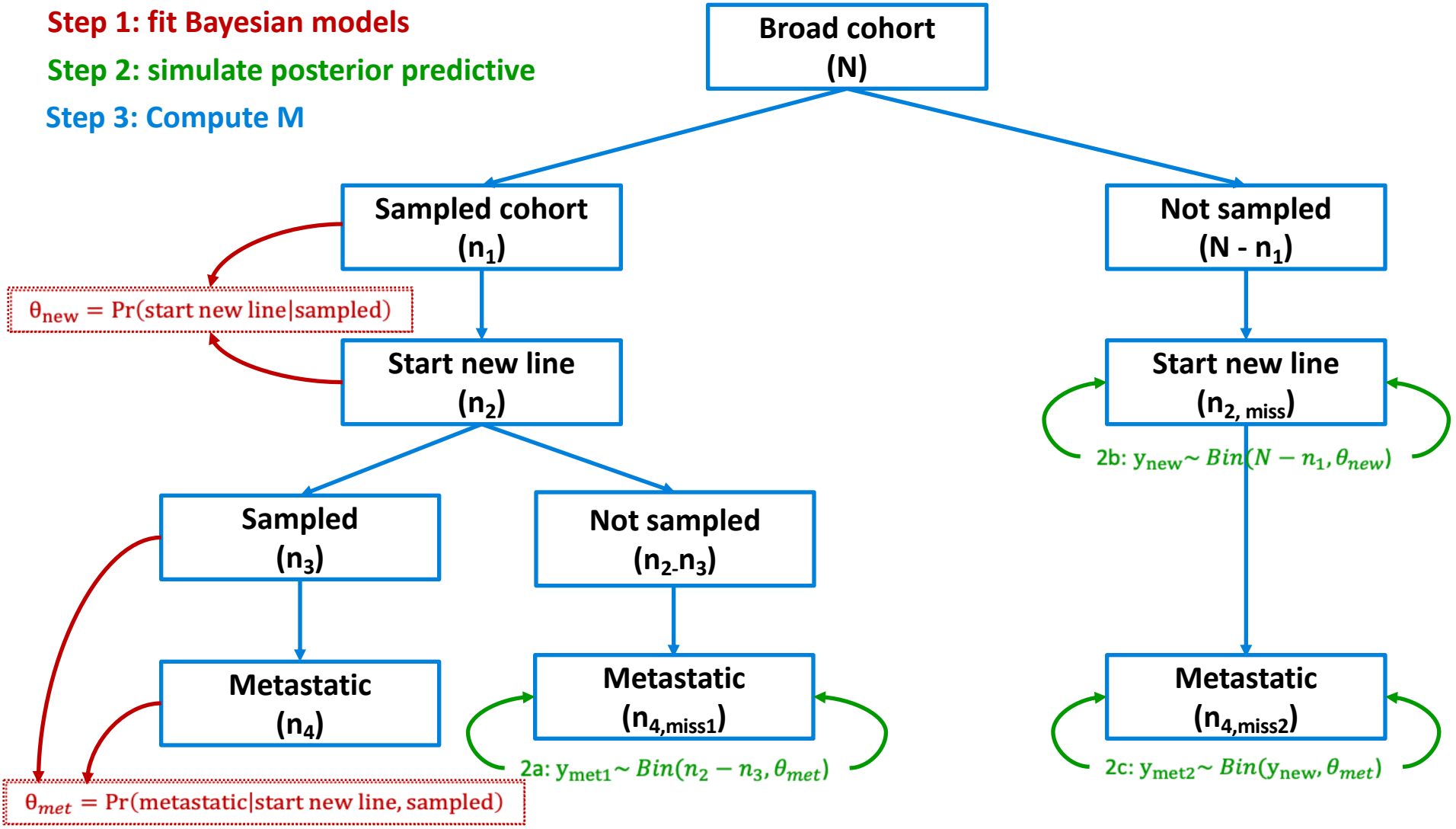
Aim is to estimate $M = n_4 + n_{4,miss1} + n_{4,miss2}$. Can we impute $n_{4,miss1}$ and $n_{4,miss2}$?

Prediction of number eligible for each site

Step 1: fit Bayesian models

Step 2: simulate posterior predictive

Step 3: Compute M



$$M = n_4 + y_{met1} + y_{met2}$$

Estimation of θ_{new} and θ_{met}

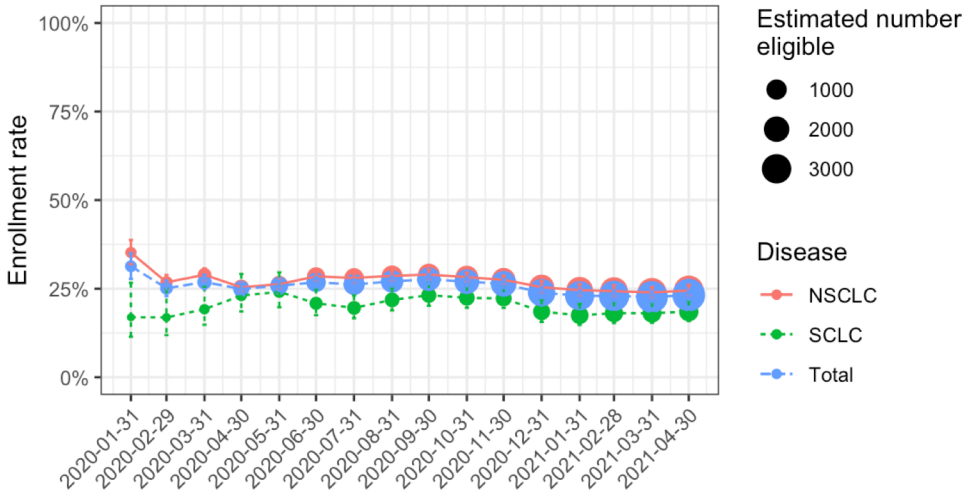
We implemented three models (Bayesian beta-binomial, frequentist, and Bayesian hierarchical), and used the hierarchical model as our primary methodology since it partially pools information across sites

$y_j \sim Bin(n_j, \theta_j)$	y_j= # metastatic n_j = # sampled with known metastatic disease (n_3)
$\theta_j = logit^{-1}(\alpha_j)$	Logistic function to model probability
$\alpha_j \sim N(\mu, \sigma^2)$	Hierarchical prior for log odds shrinks site specific estimates toward overall means
$\mu \sim N(logit(p), 0.5)$	Empirical Bayes prior where p is the proportion of the entire sample that is metastatic
$\sigma \sim N(0,2)$	Weakly informative prior truncated to be nonnegative

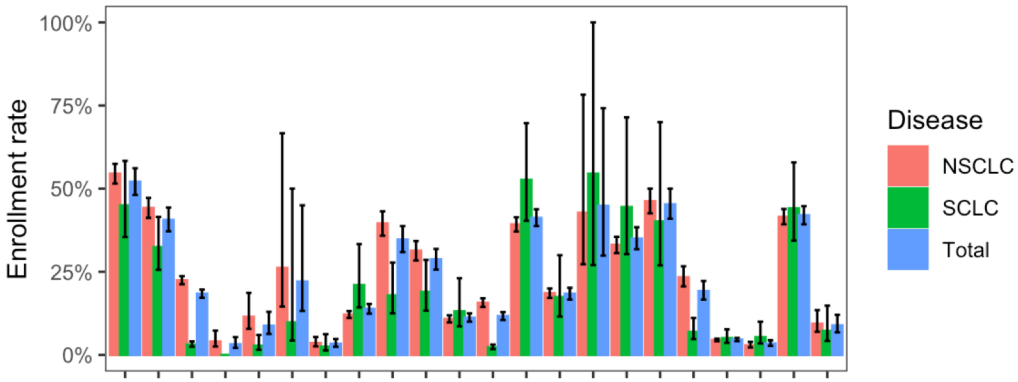
Note: j indexes site

Predicted enrollment rates from the Bayesian hierarchical model

Cumulative enrollment rate over time



Enrollment rate (April 2021) by site



The enrollment rate has been steady over time (~25%) with rates slightly higher for NSCLC than SCLC. There is considerable variation across sites and the hierarchical model improves the stability of the site specific estimates

Summary and conclusions

- PCG is a prospective pragmatic study with active collection of blood samples at three time points combined with passive collection of additional data through existing EHR systems
- We developed a methodology tailored to Flatiron Health's data collection process for estimating enrollment rates in a prospective pragmatic study
- Relatively high enrollment rates suggest that pragmatic studies with community sites are feasible
- The methodology has been used to project enrollment for PCG, including extensions to new target populations; it may also be helpful for future pragmatic studies
- A useful next step is to use the lessons learned from PCG to design and assess a randomized pragmatic trial

Doing now what patients need next