**Analysis plan: Estimating the cost of diagnostic delays for endemic mycoses**

**Project objective:** We want to model the impact of diagnostic delays on patient outcomes and cost of treatment for endemic mycoses (coccidioidomycosis, blastomycosis, histoplasmosis). We hypothesize that diagnostic delays significantly increase cost of treatment, and this model will allow us to estimate how different interventions that reduce diagnostic delays can reduce treatment costs.

**Project roles (tentative):**

* Lead author: Sam
* Lead analyst: Jason
* Mentors: Kaitlin Benedict, Mal Rajeev, Mitsuru Toda
* Advisors: Michelle Fearon Scales, Jeremy Gold, DFWED Prevention Office

**Analysis timeline:**

* July – August 2024: finalize data definitions and pull data
* September 2024: exploratory and descriptive data analyses to inform statistical methods
* October – November 2024: statistical analyses
* December 2024 – January 2025: econ/cost analyses

**Data definitions**

We will use MarketScan data from 2017 – 2022. Our patient cohort will be based on the following data definitions.

***Patient population***

* Patients with a diagnosis of cocci, histo, or blasto who are on a non-capitated plan and are continuously enrolled at least six months before and 1 year after their diagnosis date.
* Patients with a cocci, histo, or blasto diagnosis in the six months before diagnosis date will be excluded, in an attempt to capture incident diagnoses.
* For histo, also exclude anyone with a histo diagnosis code assigned by an eye care provider

***Diagnosis date***

* Date of first diagnosis using ICD-10 codes for cocci, histo, and blasto.

***Symptom onset date***

* In the three months before diagnosis date, first date of any of the compatible symptoms and signs, clinical findings, or potential misdiagnoses listed in Table 1.

***Treatment costs and indicators (Table 2)***

* Cost of treatment post diagnosis: any costs that are associated with the reason for treatment codes corresponding to infection (same as the codes used to define diagnosis date)
* Costs associated with misdiagnosis, e.g. treatment costs before diagnosis date

***Data output***

Patient-diagnosis level data on the following:

* Patient ID number
* What diagnosis (cocci histo blasto)
* Sub-code(s) identifying specific clinical form of cocci, histo, or blasto
* Date of first visit for compatible symptom/clinical finding/other respiratory diagnosis
* Date of first and last visit for endemic mycosis
* Compatible symptoms/clinical findings/other respiratory diagnosis that identified onset
* Underlying conditions: asthma, diabetes, COPD, HIV/AIDS, immune-mediated inflammatory disease, cancer
* Immunosuppression (flag that combines certain underlying conditions or immunosuppressive medications)
* Demographics: Sex, Age, Rural/non-rural, Region, State
* Costs before (up to 3 months before) and after diagnosis (on and up to 1 year after) (see Table 2)
* Insurance plan (commercial vs. Medicare advantage/supplemental)

**Statistical methods**

DAG (directed acyclic graph) of our problem. We are trying to estimate the effect of diagnostic delays on cost of treatment.

Delay to diagnosis (days)

Cost of treatment

Potential confounders (see Table 3)

*Effect we are trying to estimate*

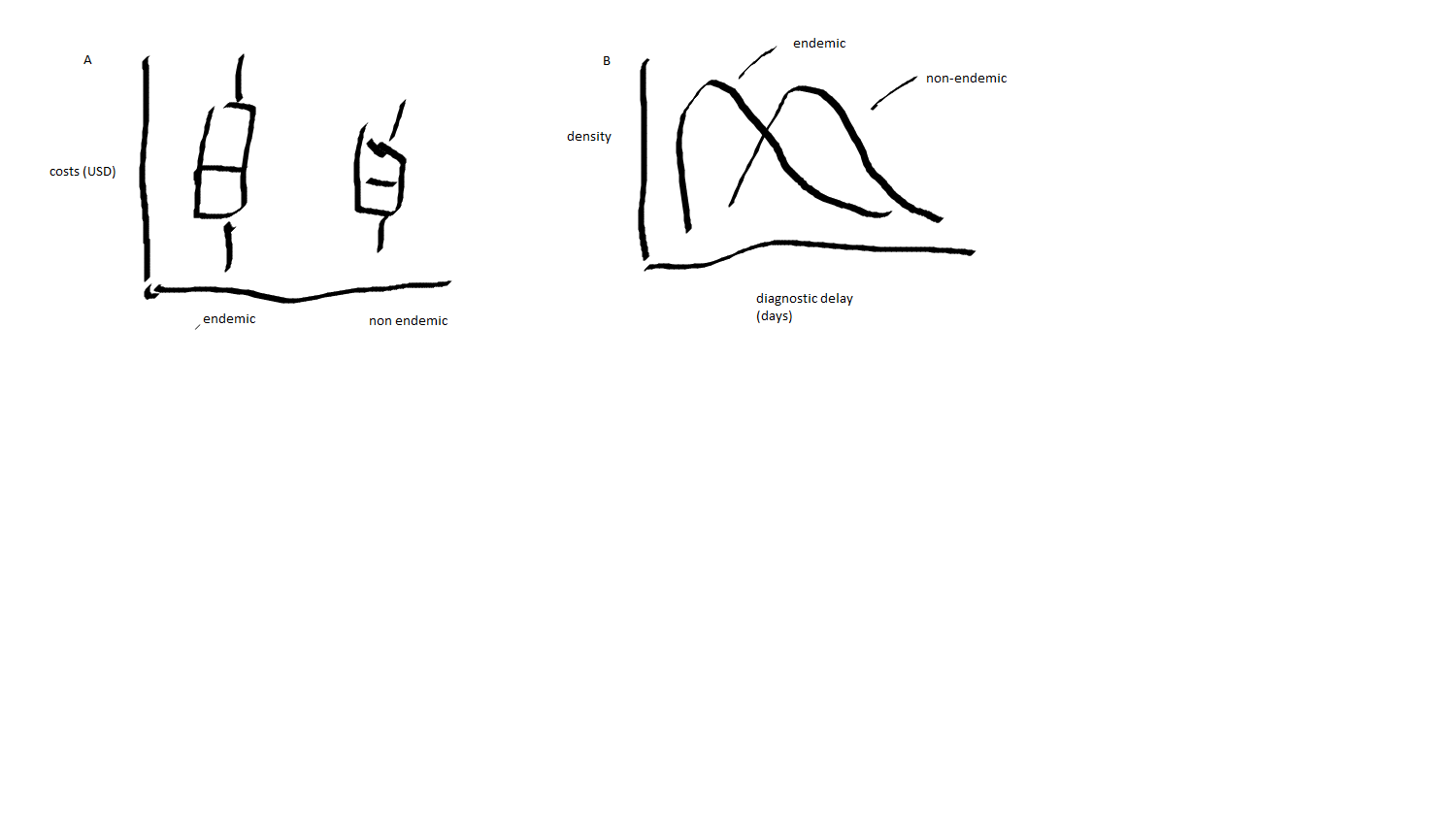
We could try:

* Multivariate with inflation-adjusted cost as the outcome (most straightforward)
* Stratified with cost as the outcome: if we identify some key factors that drive costs (immunocompromised status) or delays (endemicity)
* Generalized propensity score and matching: [Implementing matching, and weighting on GPS with continuous exposures. • CausalGPS (nsaph-software.github.io)](https://nsaph-software.github.io/CausalGPS/) to estimate the exposure-response function.
* Also this looks helpful: [CausalGPS: An R Package for Causal Inference With Continuous Exposures (arxiv.org)](https://arxiv.org/pdf/2310.00561)
* I think we should try a GAM because I’m not sure that we would expect the relationship between diagnostic delays and cost to be linear (more on [Generalized Additive Models in R · A Free Interactive Course (noamross.github.io)](https://noamross.github.io/gams-in-r-course/)**).**
* We may want to fit models to overall cost vs. costs stratified by type (outpatient vs. inpatient)

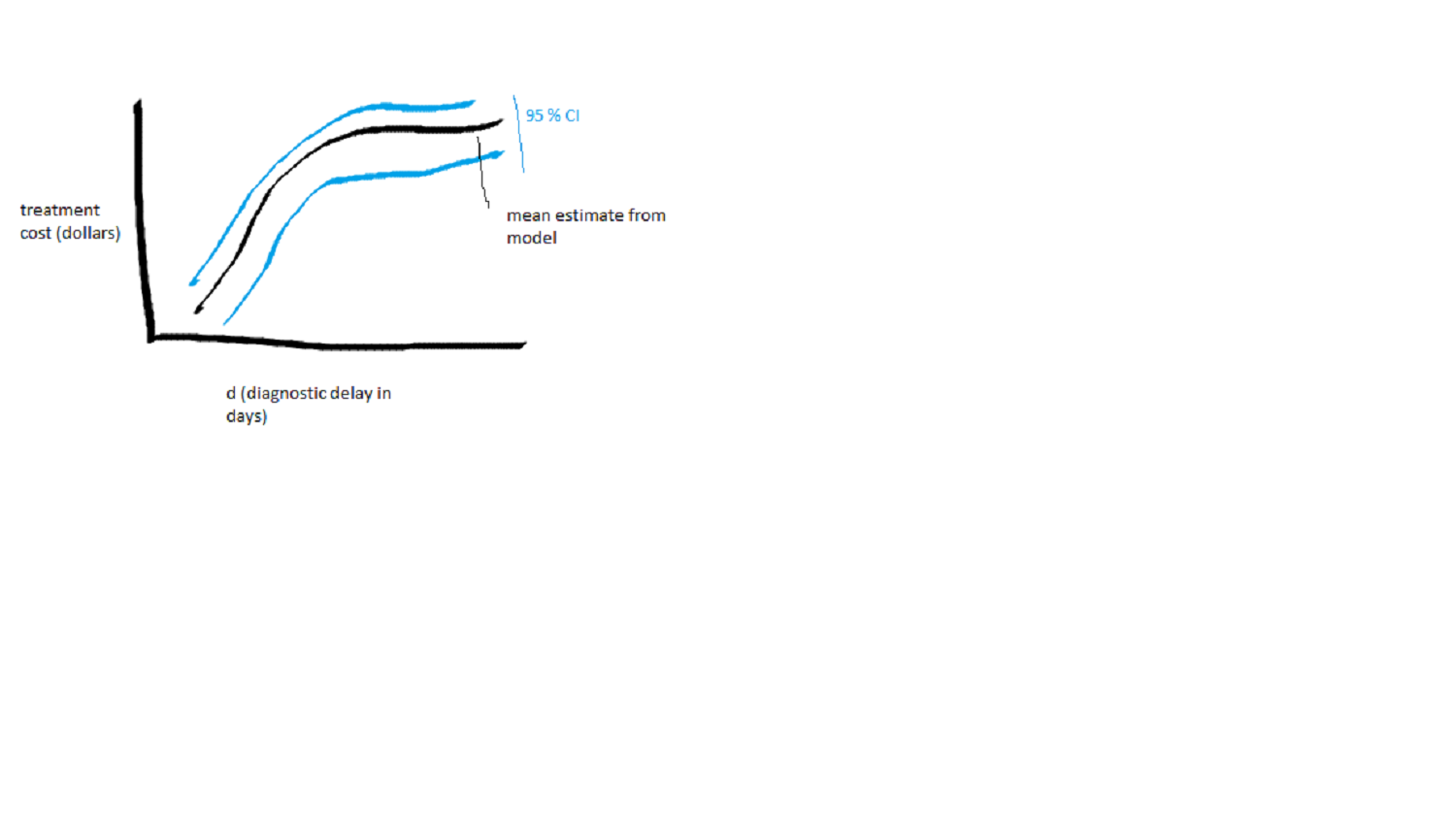
**Economic analysis methods**

**Anticipated Results**

* Descriptive analyses: overall distribution of costs and diagnostic delays for each disease
  + And also to explore potential confounders, stratified by:
    - Endemicity (mainly for cocci, but maybe also blasto)
    - Immunocompromised status
    - Disease sub-type
    - Demographics: Sex, Age, Rural/non-rural
    - Plan type?



* + *Example: Boxplots of cost and diagnosis delay distributions by endemicity status.*
* Statistical analyses: exposure response function of delays on costs (by different types of costs)



* + *Example: sketch of predicted effect of diagnostic delays on treatment costs from a GAM (non-linear)*
* Effects to report
  + Overall effects on treatment costs for unit increase in diagnosis delays
  + Potential cost savings of reducing diagnosis delays (counterfactual)
  + ??

**Potential risks and ways to mitigate them**

* Given sample sizes and potential confounders, we may not see much of a relationship between diagnostic delays and costs in these data.
  + Alternative: explore some other impacts like delays vs. hospitalization lengths or # of visits or other indicators of impact on patient outcomes
  + Look at predictors of diagnostic delays that may inform these impacts (endemicity/regional risk, demographics etc.)
* Cost data may be tricky to extrapolate beyond this dataset
  + Will need input from folks with an econ background on how we can present these results.

**Tables**

***Table 1. Possible diagnoses to flag illness onset date and pre-diagnosis costs.***

|  |  |
| --- | --- |
| **Type** | **Code (descriptive but Kaitlin has these codes)** |
| **Compatible symptoms/signs** | |  | | --- | | Abnormal weight loss | | Chest pain | | Chills, without fever | | Cough | | Dyspnea | | Erythema nodosum | | Fatigue or malaise | | Fever | | Generalized hyperhidrosis | |  | | Myalgia | | Pain in joint | |
| **Compatible clinical findings** | |  | | --- | | Acute respiratory failure | | Diseases of mediastinum | | Enlarged lymph nodes | | Hypoxemia | | Other nonspecific abnormal finding of lung field | | Pleural effusion | |
| **Other respiratory diagnoses** | |  | | --- | | Acute upper respiratory infection | | Bronchitis | | Influenza | | Lung cancer | | Other or unspecified respiratory disorder | | Pneumonia | | Unspecified acute lower respiratory infection | |

***Table 2. Cost measures and other outcomes of interest***

|  |  |
| --- | --- |
| **Timing** | **Measure** |
| 3 months before endemic mycosis diagnosis | Number of symptom-related visits |
| Costs of symptom-related visits |
| Antibiotics |
| Costs of antibiotics |
| On diagnosis date or in the year after diagnosis | Number of endemics-related outpatient visits |
| Costs of endemics-related outpatient visits |
| Hospitalized for endemic mycosis |
| Number of days hospitalized |
| Costs of hospitalizations |
| Outpatient antifungals |
| Number of days supply of outpatient antifungals |
| Costs of outpatient antifungals |

***Table 3. Potential confounders***

|  |  |  |
| --- | --- | --- |
| **Covariate** | **Impact on diagnostic delays** | **Impact on costs** |
| Immunosuppressed status/underlying conditions | Patients who are immunosuppressed (or generally sicker) may be in the healthcare system more frequently and thus receive a quicker diagnosis than healthier patients | Patients who are immunosuppressed may experience higher treatment costs due to underlying health conditions, complications, etc. |
| Endemicity/regional risk | Patients in endemic states may receive a quicker diagnosis due to awareness of risk in the healthcare setting. | Patients in endemic states might have lower costs if they have faster time to diagnosis. |
| Age | Patients who are younger or older may receive a quicker diagnosis due to lower threshold of care seeking behavior | Costs might vary differentially by age |
| Rural status | Patients in rural areas might have difficulty in accessing specialty care and thus longer times to diagnosis | Patients in rural areas might have higher costs if they have longer times to diagnosis |