Exposure and Control matching in Electronic Health Record Studies using Grouped Marching

Margaret A. Hall, MS[1], M. Daniel Brannock, MS[2], Richard A. Moffitt, PhD[1,3]

[1] Department of Hematology and Medical Oncology, Emory University School of Medicine
[2] RTI International
[3] Department of Biomedical Informatics, Emory University School of Medicine

Background

- Observational studies from EHR data are reliant on appropriate cohort creation
- EHR data only knows when patients interact with the system: simply visiting the doctor is a “risk factor” for recording a diagnosis, prescribing medication, etc.
  - Still, index dates for observational studies are typically interactions with the healthcare system
- Untested and test-negative patients are inherently different
  - COVID test-negative populations are sicker than test-positive, even after extensive matching
  - Patients with negative COVID tests may not even be under suspicion of COVID-19 infection
- Test-positive and diagnosed patients also differ (Fig. 1)

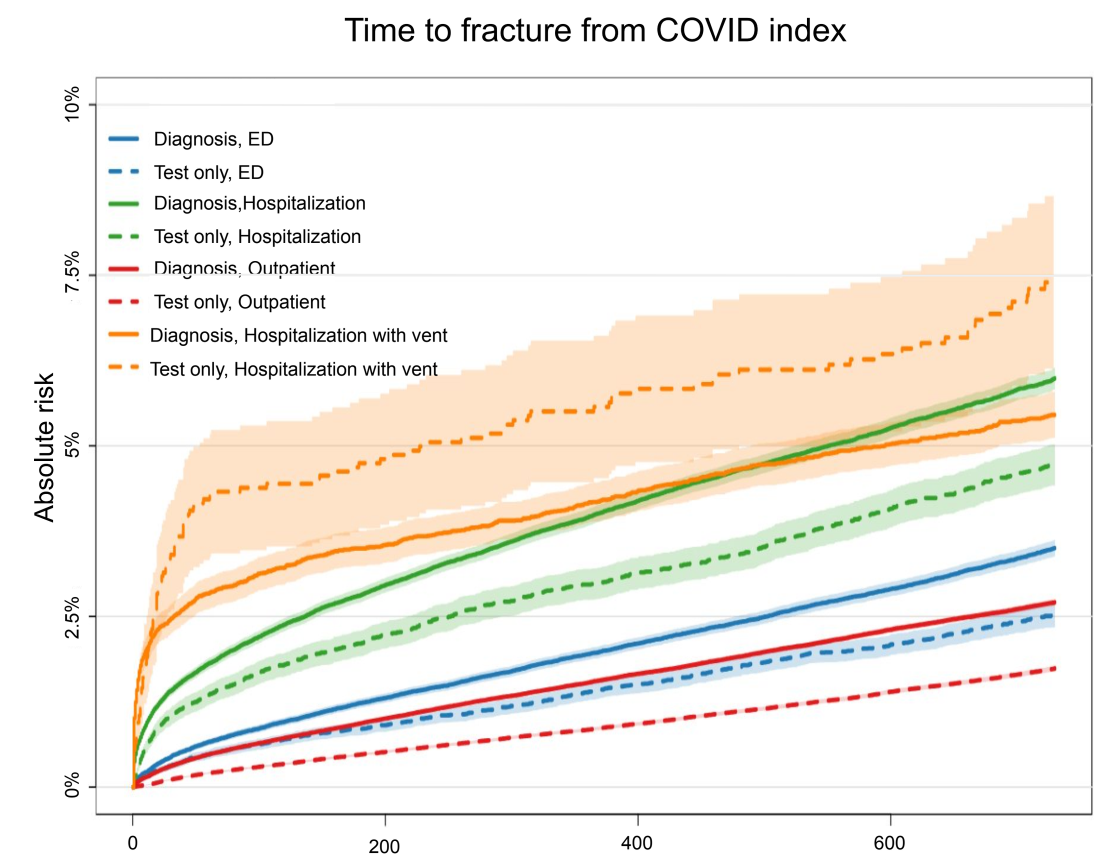

Figure 1: Risk of fracture based on severity (setting) of COVID-19 infection. Solid lines show patients with COVID-19 diagnosis, dashed lines show those with only positive test results.

Methods

Patients are *enrolled* based on an event representing “a normal day” that is unrelated to the exposure or endpoint, e.g. primary care well visit.

1. Coarsened exact matching is used to determine matching groups based on desired characteristics
2. The first exposed case in the group is randomly paired with another patient from the group
3. Both are removed from the group
4. March to next exposure in the group and repeat 2-3 until all exposures are paired

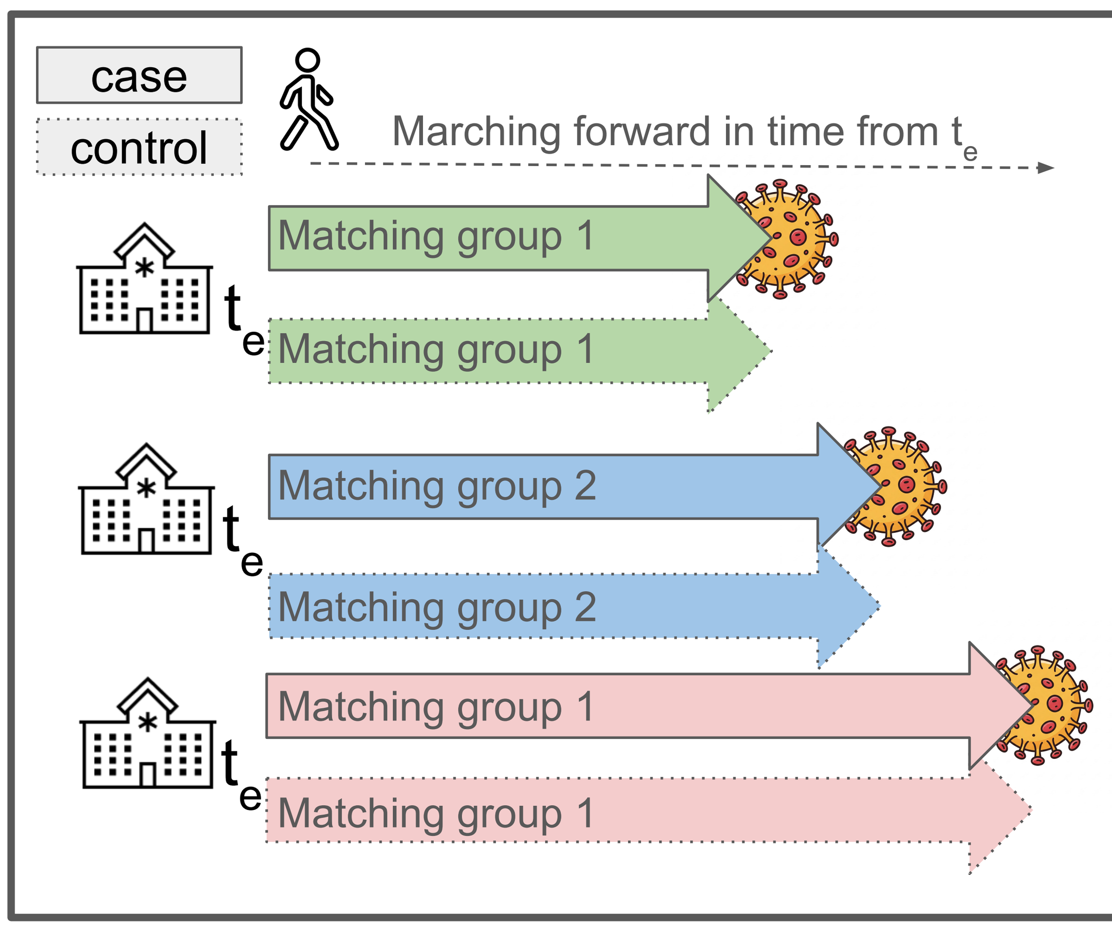

Figure 2: Pair assignment is performed chronologically from t_e. t_0 for the control is the same number of days after t_e as the paired case. Events occurring after t_0 are ignored for pair creation–a control may experience the exposure later.

Grouped marching

- t_e = time of enrollment
  - primary care/ wellness visit
- t_0 = study begins
  - Exposure event, e.g. COVID infection
  - Matched controls given t_0 index same number of days after t_e
- t_i = event occurred
- Matched control may get COVID
  - Pair is censored together if control is no longer an eligible control

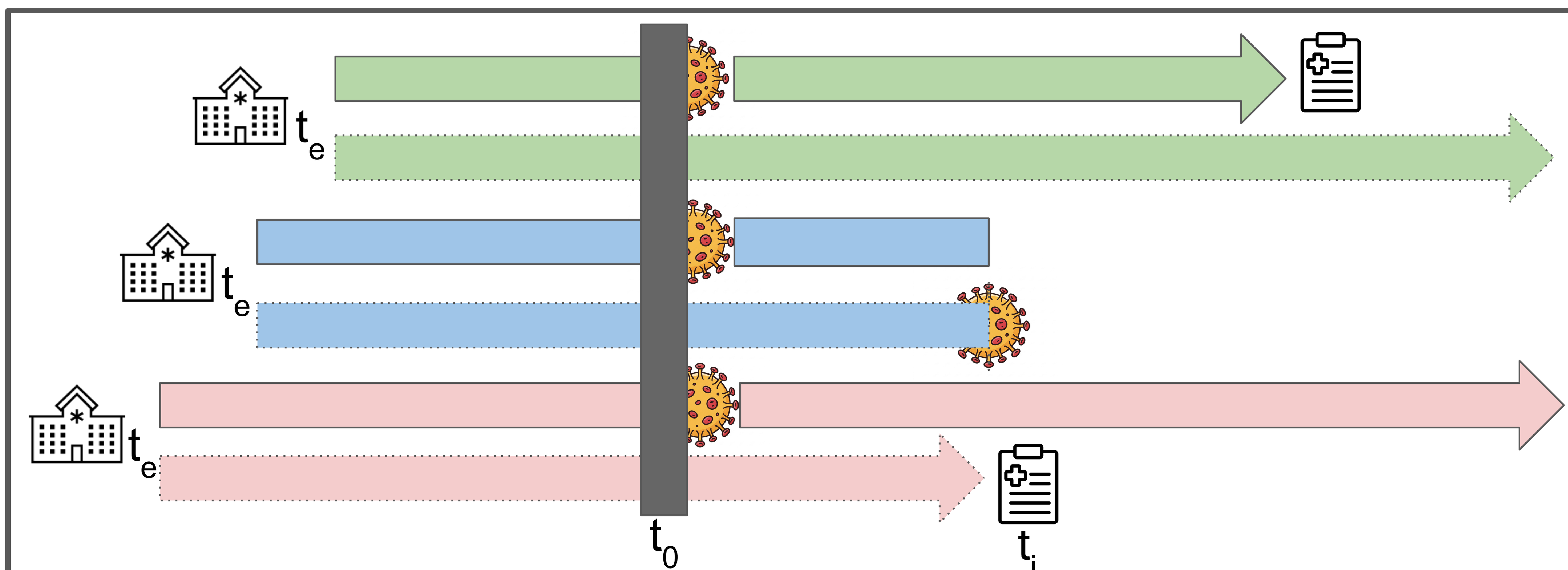

Figure 3: After pairing, t_i can be determined. A matched pairs is censored if the control is no longer eligible as a control due to experiencing the exposure event.

Results

- Enrollment is COVID-19 infection, exposure is a reinfection, endpoint is Long COVID (LC)
  - 174,917 eligible matched patients
- Negative controls that we believe should not be correlated with exposure (non-pathological fracture, pap smear) have no difference in risk
- LC risk is significantly increased after reinfection[1]

Primary	
Reinfection -> LC	1.32(1.29 to 1.35)
Split by Age	
Reinfection -> LC, 18-25	1.40(1.22 to 1.58)
Reinfection -> LC, 26-35	1.37(1.25 to 1.48)
Reinfection -> LC, 36-50	1.39(1.32 to 1.46)
Reinfection -> LC, 51-65	1.30(1.23 to 1.36)
Reinfection -> LC, 65+	1.23(1.17 to 1.29)
Vaccination Status	
Reinfection -> LC, vax_before_and_between	1.17(1.00 to 1.35)
Reinfection -> LC, vax_just_before_first	1.34(1.20 to 1.49)
Reinfection -> LC, vax_just_between_infs	1.22(1.11 to 1.34)
Reinfection -> LC, vax_none	1.47(1.36 to 1.58)
Negative Control	
Reinfection -> Fracture	1.06(0.97 to 1.14)
Reinfection -> Pap Smear	1.12(0.97 to 1.34)

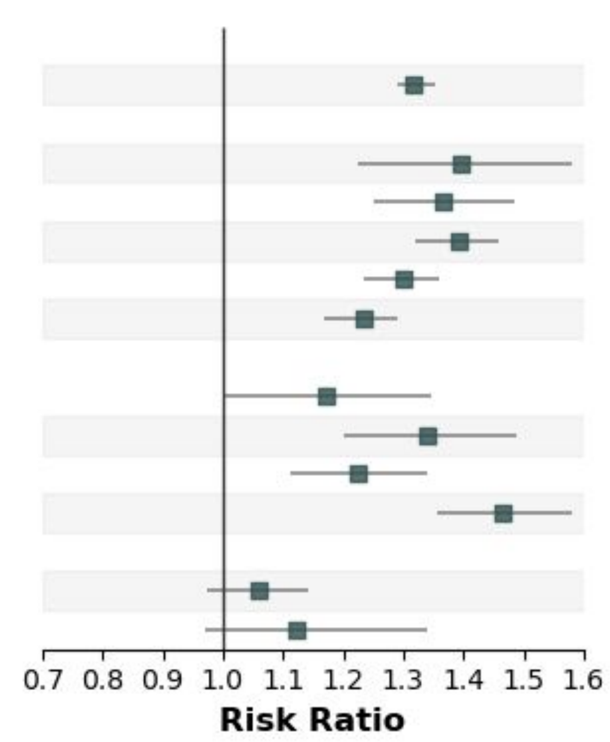

Figure 4: After reinfection, risk of fracture and pap smear are non-significant. Risk of long covid is significantly increased. This is true of all age groups and most vaccination groups.

Conclusion and future work

- Grouped Marching removes bias between case and control in unrelated outcomes.
- Only cases have an encounter on day 0, which may increase all observed events.
- Simulation studies underway.

References

1. Brannock, M. Daniel, Emily Hadley, Alexander Preiss, Megan L. Fitzgerald, Nita Jain, Emily Taylor, Andrew Wylam, Yun J. Yoo, Elaine Hill, and Richard A. Moffitt. "Incidence of Long COVID Following Reinfection with COVID-19." medRxiv (2025): 2025-08.