Defining an appropriate healthy reference cohort within a network of children's hospital health systems (PEDSnet)

Amy J. Goodwin Davies, PhD[1], Michelle R. Denburg, MD, MSCE[1], H. Timothy Bunnell, PhD[3], Peter E. F. Camacho[1], L. Charles Bailey, MD, PhD[1], Melody M. Kitzmiller[2], Rui Xiao, PhD[1], Hanieh Razzaghi, MPH[1]

[1]Applied Clinical Research Center, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[2]Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH, USA
[3]Nemours Biomedical Research, Wilmington, DE, USA

Abstract *Studies which investigate outcomes for a case cohort of patients will often aim to compare these outcomes to the general "healthy" population. Cohorts of "healthy" populations, particularly among children, are also important to define baseline care practices. This poses a challenge for studies conducted within a network of hospital health systems, due to an over-representation of patients seeking specialty care. We describe a rule-based approach to defining a "healthy" reference cohort for a population-based retrospective cohort study investigating outcomes for a rare disease within PEDSnet. PEDSnet is a network of 7 children's hospital health systems which contains data for more than 6 million children. PEDSnet uses an expanded version of the OMOP CDM[1].*

Research Category Clinical characterization / Other: Cohort definition

Background For a population-based retrospective cohort study investigating the incidence of skeletal outcomes in a cohort of pediatric patients with glomerular disease[2], we defined a "healthy" reference cohort (henceforth HR cohort). The study is reported elsewhere[3]; the focus of this paper is the generalizable approach to defining a HR cohort to compare to a rare disease cohort (henceforth RD cohort) using the OMOP CDM. Rather than defining a control cohort with maximal similarity to RD cohort aside from the rare disease, our goal was to define a cohort which is representative of the general pediatric population, which is mostly healthy.

Methods Our general approach to the HR cohort was to restrict to children receiving ongoing primary care within PEDSnet and to exclude children with progressive disease and those for whom most encounters with the network are related to the outcome under study. Given this focus on primary care, we restricted the HR cohort to institutions within PEDSnet with a primary care network. In the following sections, we discuss the motivation and implementation of the following inclusion ([INCL]) and exclusion ([EXCL]) criteria: (1) [INCL] Follow-up of at least one year, (2) [EXCL] Most encounters related to outcome under study, (3) [INCL] Ongoing primary care, (4) [EXCL] Condition associated with deteriorating health or member of the RD cohort. Criteria (1) and (2) were applied to the RD cohort as well as the HR cohort.
[INCL] Follow-up of at least one year: A follow-up requirement of at least two outpatient encounters separated by at least one year was implemented for both the RD cohort and the HR cohort. Motivations were as follows: (i) Patients who interact with the network over an extended period of time are more likely to be seeking medical care within the network and will therefore have more complete follow-up, compared to patients who only interact with the network for a shorter period of time. (ii) We anticipate that some children within the PEDSnet population will interact with the network for a short period of time exclusively for treatment of an acute condition or the outcome of a particular study (e.g., bone fractures). Including these patients inflates the incidence of the outcome as these patients would contribute a small amount of person-time per outcome event. Given patients with a rare disease generally receive continued care in tertiary care centers, this overestimation would affect the HR cohort to a greater extent than the RD cohort.
[EXCL] Most encounters related to outcome under study: We did not want to include patients who were interacting predominantly for treatment related to the outcome under study. For skeletal outcomes, we implemented this by excluding patients if they had more orthopedic visits, defined as an encounter with a clinic or provider with an orthopedic specialty, than visits of any other specialty (PEDSnet's CDM expands the OMOP CDM to include specialty information for clinics as well as providers). The motivation here was to remove patients with recurrent orthopedic issues as they are not representative of the general "healthy" population and would inflate skeletal outcome incidence rates.
[INCL] Ongoing primary care: In addition to primary and secondary care, institutions within PEDSnet provide tertiary and quaternary care so we expect that illness, injury, and other medical conditions are over-represented in the entire PEDSnet population compared to the general "healthy" population. We ameliorated this risk by requiring that patients in the HR cohort receive ongoing primary care. This was implemented as two or more outpatient visits with a clinic or provider with a general practice specialty and at least one general practice outpatient encounter every 18 months across a patient's follow-up period. A secondary motivation is that patients who exclusively interact with the network for specialty care are likely to seek other medical care outside the network, including treatment for the outcome under study.
[EXCL] Condition associated with deteriorating health or member of RD cohort: As the HR cohort should capture mostly "healthy" children, we excluded patients with "progressive" disease; conditions associated with deteriorating health or reduced life expectancy[4]. This exclusion was implemented by excluding

patients with any diagnosis flagged as "progressive" in a Pediatric Medical Complexity Algorithm[4]. We additionally excluded patients who met the cohort definition for the RD cohort (glomerular disease).

	Criterion	Count (% PEDSnet restricted to sites with primary care network)
1	[INCL] Follow-up of at least one year	1,633,940 (46.5%)
2	[EXCL] Most encounters related to outcome under study	1,488,443 (42.4%)
3	[INCL] Ongoing primary care	635,001 (18.1%)
4	[EXCL] Progressive condition or member of RD cohort	553,624 (15.8)

We considered an approach which samples from the HR cohort matching on distributions in the RD cohort (such as sex, years follow-up, and age at first visit). We ultimately decided against this approach for several reasons: (i) There are advantages of defining a HR cohort without closely matching to the RD cohort. If the HR cohort is not contingent upon another cohort in this way, this allows comparisons from multiple cohorts to the same reference cohort and furthermore allows us to ask epidemiological questions of the HR cohort separately. (ii) In some cases, we expect distinct distributions for demographic attributes, years follow-up, and age at first visit, among a RD cohort and a HR cohort. This raises the question of whether by selecting for patients with matching distributions, we would sample an unrepresentative HR cohort. (iii) A matched approach, if sampling without replacement, limits our HR cohort size to the minimum number of matches within any category of variables (i.e. if there were 3 matches in the smallest sex×follow-up×age category, the match would be limited to 3:1). Given the reasons above and considering that the planned incidence analyses for our specific use-case adjust for sex, age, and follow-up we decided to adopt a non-matched approach which compares the RD cohort to all patients in the HR cohort. A non-matched approach may lead to imbalances in age, sex, race, and follow-up. Depending on the analytic methods of the study, a variety of approaches could be used to sample or weight patients according to demographic factors and follow-up time.

Results We are investigating measures of face validity, with an initial focus on the skeletal outcomes under study. The incidence rate of bone fracture per 10,000 person-years within each year of age aligns with findings in the literature for another primary care cohort: Cooper et al[5] investigate rates of fracture within general practice research database with similar incidence curves to those in our HR cohort (Figure 1). They find a peak of ≈160 per 10,000 person-years at age 11 for girls and a peak of ≈280 at age 14 for boys, compared to 138 at age 11 for girls and 270 at age 13 for boys in our HR cohort. Turning to other clinical characteristics, the most frequent diagnoses provide preliminary support that this cohort is representative of the general "healthy" population: pharyngitis, upper respiratory infection, otitis media, cough, and allergic rhinitis, conditions we expect to find in children without medical complexity.

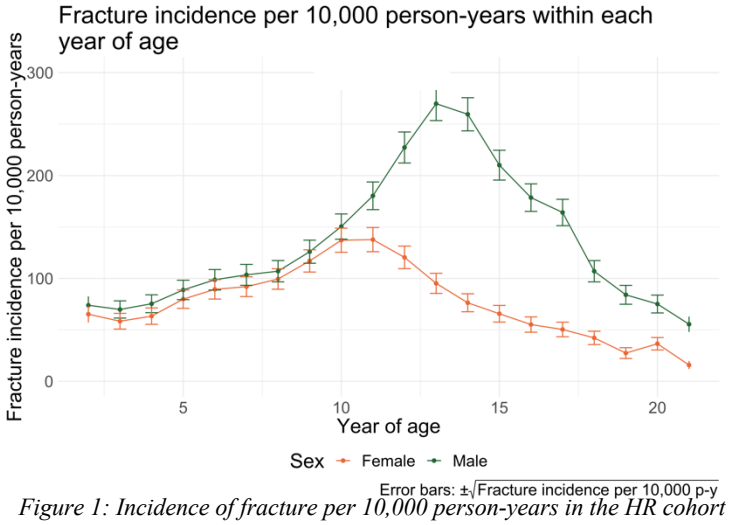

Figure 1: Incidence of fracture per 10,000 person-years in the HR cohort

Discussion/Conclusion We have developed a generalizable approach to defining a HR cohort, implemented using the OMOP CDM. In our approach, the HR cohort is not closely tied to a specific cohort, which has the advantage of a single reference cohort being used for comparisons across multiple cohorts and the ability to investigate epidemiological questions of the HR cohort independently. Ongoing analyses, initially focused on the skeletal outcomes under study, provide face validity. Additional work is required to validate that our approach defines an HR cohort, which is representative of the general "healthy" population both for the skeletal outcomes study described here and across additional future studies.

References

1. Forrest C, Margolis P, Bailey LC, Marsolo K, Del Beccaro M, …, Kahn M. PEDSnet: a national pediatric learning health system. J Am Med Inform Assoc. 2014; 21(4):602-606.
2. Denburg MR, Razzaghi H, Bailey LC, Soranno DE, Pollack AH, ..., Flynn JT. Using electronic health record data to rapidly identify children with glomerular disease for clinical research. J Am Soc Nephrol. 2019; 30(12):2427-2435.
3. Goodwin Davies AJ, Razzaghi H, Bailey LC, Meloni S, Utidjian LH, …, Denburg M. Skeletal outcomes in children and young adults with glomerular disorders. Paper presented at: Pediatric Academic Societies/American Society of Pediatric Nephrology Basic and Clinical Research Platform Presentations hosted by the Pediatric Nephrology Research Consortium; 2020 May 5; Virtual.
4. Simon TD, Cawthon ML, Stanford S, Popalisky J, Lyons D, ..., Mangione-Smith R. Pediatric medical complexity algorithm: a new method to stratify children by medical complexity. Pediatrics. 2014; 133(6):e1647-e1654.
5. Cooper C, Dennison EM, Leufkens HG, Bishop N, van Staa TP. Epidemiology of childhood fractures in Britain: a study using the general practice research database. J Bone Miner Res. 2004; 19(12):1976-1981.