



American Society of Nephrology
1401 H St NW, Suite 900
Washington, DC 20005
Phone: 202-640-4660 | Fax 202-637-9793
vramsey@kidney360.org

The information on this cover page is based on the most recent submission data from the authors. It may vary from the final published article. Any fields remaining blank are not applicable for this manuscript.

Article Type: Original Investigation

Validating a computable phenotype for nephrotic syndrome in children and adults using PCORnet® data

DOI: 10.34067/KID.0002892021

Andrea Oliverio, Dorota Marchel, Jonathan Troost, Isabelle Ayoub, Salem Almaani, Jessica Greco, Cheryl Tran, Michelle Denburg, Michael Matheny, Chad Dorn, Susan Massengill, Hailey Desmond, Debbie Gipson, and Laura Mariani

Key Points:

*A computable phenotype combines routinely collected data elements from the EHR with logic elements to identify a condition of interest

*This validated computable phenotype has strong classification characteristics to identify individuals with primary nephrotic syndrome

*This computable phenotype for primary nephrotic syndrome can facilitate future research of these rare diseases

Abstract:

Background: Primary nephrotic syndromes are rare diseases which impedes adequate sample size for observational patient-oriented research and clinical trial enrollment. A computable phenotype may be powerful in identifying patients with these diseases for research across multiple institutions. **Methods:** A comprehensive algorithm of inclusion and exclusion ICD-9 and ICD-10 codes to identify patients with primary nephrotic syndrome was developed. The algorithm was executed against the PCORnet® CDM at 3 institutions from Jan 1, 2009 to Jan 1, 2018, where a random selection of 50 cases and 50 non-cases (individuals not meeting case criteria seen within the same calendar year and within five years of age of a case) were reviewed by a nephrologist, for a total of 150 cases and 150 non-cases reviewed. The classification accuracy (sensitivity, specificity, positive and negative predictive value, F1 score) of the computable phenotype was determined. **Results:** The algorithm identified a total of 2,708 patients with nephrotic syndrome from 4,305,092 distinct patients in the CDM at all sites from 2009-2018. For all sites, the sensitivity, specificity, and area under the curve of the algorithm were 99% (95% CI: 97-99%), 79% (95% CI: 74-85%), and 0.9 (0.84-0.97), respectively. The most common causes of false positive classification were secondary FSGS (9/39) and lupus nephritis (9/39). **Conclusion:** This computable phenotype had good classification in identifying both children and adults with primary nephrotic syndrome utilizing only ICD-9 and ICD-10 codes, which are available across institutions in the United States. This may facilitate future screening and enrollment for research studies and enable comparative effectiveness research. Further refinements to the algorithm including use of laboratory data or addition of natural language processing may help better distinguish primary and secondary causes of nephrotic syndrome.

Disclosures: L. Mariani reports the following: Consultancy Agreements: Reata Pharmaceuticals, CKD Advisory Committee; Travere Therapeutics, Advisory Board; Calliditas Therapeutics, Advisory Board; Research Funding: Boehringer Ingelheim; Honoraria: Reata Pharmaceuticals, CKD Advisory Committee; Travere Therapeutics, Advisory Board; Calliditas Therapeutics, Advisory Board; ASN BRCU; Scientific Advisor or Membership: Travere Therapeutics; Calliditas Therapeutics; Reata Pharmaceuticals. S. Almaani reports the following: Consultancy Agreements: Aurinia Pharmaceuticals; Kezar Life Sciences; Research Funding: Gilead Sciences; Scientific Advisor or Membership: Clinical Nephrology - Editorial Board; Speakers Bureau: Aurinia Pharmaceuticals. I. Ayoub reports the following: Honoraria: ACR; Scientific Advisor or Membership: Lupus Foundation of America (advisory board); Journal of Clinical Nephrology (Editorial Board). M. Denburg reports the following: Consultancy Agreements: Trisalus Life Sciences (spouse); Ownership Interest: Precision Guided Interventions LLC (spouse); In-Bore LLC (spouse); Research Funding: Mallinckrodt; Scientific Advisor or Membership: NKF Delaware Valley Medical Advisory Board; Trisalus Life Sciences Scientific Advisory Board (spouse); Other Interests/Relationships: American Society of Pediatric Nephrology Research and Program Committees; National Kidney Foundation Pediatric Education Planning Committee. H. Desmond reports the following: Research Funding: I receive a percentage of my salary from the University of Michigan but funded by the following groups: Boehringer Ingelheim. D. Gipson reports the following: Consultancy Agreements: Through University of Michigan: Vertex Pharmaceuticals; AstraZeneca; Boehringer Ingelheim; Roche/Genentech; Research Funding: Through University of Michigan: National Institutes of Health, Food and Drug Administration, Centers for Disease Control; Travere ; Novartis; Goldfinch Bio; Boehringer Ingelheim; Atrium Health Medical Foundation; Reata; Scientific Advisor or

Membership: Member American Society of Nephrology; International Society of Pediatric Nephrology; American Society of Pediatric Nephrology. S. Massengill reports the following: Consultancy Agreements: Guidepoint Group; Scientific Advisor or Membership: Editorial Board of online Glomerular Disease Journal (Kargers Publishing). M. Matheny reports the following: Consultancy Agreements: NIH-VA-DoD Pain Management Grant Consortium (PMC3); Scientific Advisor or Membership: SMRB Study Section, VA HSR&D, Informatics & Methods Section; Steering Committee - Indianapolis VA HSR&D COIN Center; Steering Committee - VA HSR&D VIREC. C. Tran reports the following: Scientific Advisor or Membership: Frontiers in Pediatrics - Review Editor on the Editorial Board of Pediatric Nephrology. The remaining authors have nothing to disclose.

Funding: NephCure Kidney Network for Patients with Nephrotic Syndrome: Andrea L. Oliverio, Michelle R. Denburg, Laura H Mariani, PPRN-1306-04903; HHS | NIH | National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK): Andrea L. Oliverio, 5T32DK007378-39 (June 2018-2019),5KL2TR002241-04 (March 2020-Sept 2020),K23 DK123413 (Sept 2020-); HHS | NIH | National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK): Laura H Mariani, K23 DK115891

Author Contributions: Andrea Oliverio: Data curation; Formal analysis; Project administration; Validation; Writing - original draft; Writing - review and editing Dorota Marchel: Data curation; Validation; Writing - review and editing Jonathan Troost: Formal analysis; Methodology; Writing - review and editing Isabelle Ayoub: Data curation; Validation; Writing - review and editing Salem Almaani: Data curation; Validation; Writing - review and editing Jessica Greco: Data curation; Validation; Writing - review and editing Cheryl Tran: Data curation; Validation; Writing - review and editing Michelle Denburg: Conceptualization; Funding acquisition; Methodology; Writing - review and editing Michael Matheny: Conceptualization; Methodology; Software; Writing - review and editing Chad Dorn: Methodology; Software; Writing - review and editing Susan Massengill: Data curation; Writing - review and editing Hailey Desmond: Data curation; Project administration; Validation; Writing - review and editing Debbie Gipson: Conceptualization; Data curation; Formal analysis; Methodology; Supervision; Writing - review and editing Laura Mariani: Conceptualization; Data curation; Formal analysis; Funding acquisition; Resources; Supervision; Validation; Visualization; Writing - review and editing

Clinical Trials Registration: No

Registration Number:

Registration Date:

How to Cite this article: Andrea Oliverio, Dorota Marchel, Jonathan Troost, Isabelle Ayoub, Salem Almaani, Jessica Greco, Cheryl Tran, Michelle Denburg, Michael Matheny, Chad Dorn, Susan Massengill, Hailey Desmond, Debbie Gipson, and Laura Mariani, Validating a computable phenotype for nephrotic syndrome in children and adults using PCORnet® data, Kidney360, Publish Ahead of Print, 10.34067/KID.0002892021

Validating a computable phenotype for nephrotic syndrome in children and adults using PCORnet® data

Andrea L. Oliverio,¹ Dorota Marchel,² Jonathan P. Troost,³ Isabelle Ayoub,⁴ Salem Almaani,⁴ Jessica Greco,⁴ Cheryl L. Tran,⁵ Michelle R. Denburg,⁶ Michael Matheny,^{7,8} Chad Dorn⁸ Susan F. Massengill,⁹ Hailey Desmond,² Debbie S. Gipson,² Laura H. Mariani¹

Affiliations:

¹ Department of Internal Medicine – Nephrology, University of Michigan, Ann Arbor MI, USA.

² Department of Pediatrics, Division of Nephrology, University of Michigan, C S Mott Children's Hospital, Ann Arbor, Michigan, USA

³ Michigan Institute for Clinical and Health Research, University of Michigan, Ann Arbor MI, USA

⁴ Department of Internal Medicine, Division of Nephrology, The Ohio State University Wexner Medical Center, Columbus, OH, USA

⁵ Division of Pediatric Nephrology, Department of Pediatrics, Mayo Clinic, Rochester, MN, United States

⁶ Division of Nephrology, The Children's Hospital of Philadelphia, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

⁷ Department of Veterans Affairs, Vanderbilt University, Nashville, US

⁸ Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, US

⁹ Levine Children's Hospital at Atrium Health, Charlotte, NC, United States

Corresponding Author: Laura H. Mariani, University of Michigan, 1500 E. Medical Center Drive, Ann Arbor, MI, 48109.

Key points:

- A computable phenotype combines routinely collected data elements from the EHR with logic elements to identify a condition of interest
- This validated computable phenotype has strong classification characteristics to identify individuals with primary nephrotic syndrome
- This computable phenotype for primary nephrotic syndrome can facilitate future research of these rare diseases

Abstract:

Background: Primary nephrotic syndromes are rare diseases which impedes adequate sample size for observational patient-oriented research and clinical trial enrollment. A computable phenotype may be powerful in identifying patients with these diseases for research across multiple institutions.

Methods: A comprehensive algorithm of inclusion and exclusion ICD-9 and ICD-10 codes to identify patients with primary nephrotic syndrome was developed. The algorithm was executed against the PCORnet® CDM at 3 institutions from Jan 1, 2009 to Jan 1, 2018, where a random selection of 50 cases and 50 non-cases (individuals not meeting case criteria seen within the same calendar year and within five years of age of a case) were reviewed by a nephrologist, for a total of 150 cases and 150 non-cases reviewed. The classification accuracy (sensitivity, specificity, positive and negative predictive value, F1 score) of the computable phenotype was determined.

Results: The algorithm identified a total of 2,708 patients with nephrotic syndrome from 4,305,092 distinct patients in the CDM at all sites from 2009-2018. For all sites, the sensitivity, specificity, and area under the curve of the algorithm were 99% (95% CI: 97-99%), 79% (95%

CI: 74-85%), and 0.9 (0.84-0.97), respectively. The most common causes of false positive classification were secondary FSGS (9/39) and lupus nephritis (9/39).

Conclusion: This computable phenotype had good classification in identifying both children and adults with primary nephrotic syndrome utilizing only ICD-9 and ICD-10 codes, which are available across institutions in the United States. This may facilitate future screening and enrollment for research studies and enable comparative effectiveness research. Further refinements to the algorithm including use of laboratory data or addition of natural language processing may help better distinguish primary and secondary causes of nephrotic syndrome.

Introduction:

Nephrotic syndrome is defined as the confluence of heavy proteinuria (>3.5 grams per day), hypoalbuminemia, and edema.¹ This syndrome is further classified by kidney histopathologic findings, including minimal change disease (MCD), focal segmental glomerulosclerosis (FSGS), and membranous nephropathy (MN). It is well-recognized that this syndrome and these histologic findings can occur secondarily to systemic disorders, viruses, malignancies or medications, or may occur in isolation, termed primary nephrotic syndrome. In addition, even when excluding presumed secondary causes of nephrotic syndrome, those remaining have heterogeneous disease features spanning their molecular mechanisms,^{2,3} histopathology,⁴ responsiveness to treatment, and outcomes.^{5,6}

Primary nephrotic syndromes are rare diseases: MCD is estimated to affect 0.23-15.6/100,000 children and 0.06/100,000 adults per year, FSGS 0.2-1.1/100,000/year, and MN 1.2/100,000/year.⁷ Their rarity and heterogeneity contribute to sample size limitations, thus performing rigorous comparative effectiveness and outcomes research is challenging. However, this work is important and necessary for patients and their caregivers, who prioritize outcomes of kidney health and survival while articulating that their lived experiences, symptoms, and functionality remain important in the assessment of successful disease management.⁸

A computable phenotype is a clinical condition, characteristic, or set of clinical features that can be determined solely from the data in electronic health records (EHR) using an algorithm and ancillary data sources without requiring chart review or interpretation by a clinician.⁹ Routinely collected data elements from the EHR combined with logic elements such as AND, OR, and IF are used to identify a clinical condition or characteristic of interest from a given data source. Applications of machine learning models with natural language processing can also develop probabilistic predictions for phenotyping, which can identify cases of disease by supervised learning from example cases.¹⁰ Valid computable phenotype algorithms for rare

conditions such as nephrotic syndrome could identify large, robust and representative retrospective cohorts that otherwise would be cumbersome to recruit prospectively. Additionally, this could improve feasibility assessment and site selection for clinical trials. Phenotypes can be replicated across data sources and healthcare organizations to generate consistent cohort identification which can streamline enrollment in registries and facilitate comparative effectiveness research.

PCORnet®, the National Patient-Centered Clinical Research Network, incorporates data from millions of patients across 9 clinical research networks and two health plan research networks including more than 300 hospitals. Data are maintained locally in participating health systems and have a shared structure (Common Data Model; CDM) that facilitates application of a computable phenotype.¹¹ Computable phenotypes developed and utilized for research within PCORnet® include those for type II DM^{12,13} and heart failure.¹⁴ Other examples of computable phenotypes in action include the Chronic Conditions Data Warehouse,¹⁵ which utilizes administrative claims from Centers for Medicare and Medicaid Services, and the Clinical Classifications Software developed as part of the Healthcare Cost and Utilization Project.¹⁶ Prior computable phenotypes developed for glomerular diseases have been shown to be reliable and valid, but focused on the pediatric population and utilized data elements which are not universally available including provider type and SNOMED-CT (Systematized Nomenclature of Medicine Clinical Terms) codes.¹⁷

We sought to develop and validate a computable phenotype to identify both adult and pediatric patients with primary nephrotic syndrome, utilizing universally available data elements to reliably capture prevalent patients in the PCORnet® CDM.

Methods:

Case definition and computable phenotype development

Electronic medical record data was originally extracted for 12,233 patients seen for glomerular disease and proteinuria at the University of Michigan from 2009-2014 with a broad range of diagnoses. A comprehensive list of Intelligent Medical Object (IMO) codes from patient problem lists and International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes was generated from this population, and the list of codes was reviewed by three investigators. The codes were cross-referenced with biopsy-proven kidney diseases attributed to primary nephrotic syndrome, or for patients without biopsy the clinician's diagnosis in the chart was reviewed. Preliminary inclusion codes were established for nephrotic syndrome utilizing those that are largely defined by the pathologic lesion determined by kidney biopsy. However, these codes do not specify if the nephrotic syndrome is idiopathic or secondary to another underlying condition, thus exclusion codes were selected for conditions known to cause secondary nephrotic syndrome (e.g., diabetes, lupus, viral illnesses). Inclusion and exclusion criteria were refined based on these codes, existing literature, and clinical rationale, and then the codes and an initial algorithm were iteratively revised to optimize sensitivity and specificity in two health systems. Subsequently, as part of the Renal Collaborative Research Group within PCORnet®, these codes were reviewed by a team of clinician scientists in 2018. Tools for mapping ICD-9-CM to ICD-10-CM codes were utilized so that ICD-10-CM codes could be added to facilitate searching beyond 2015.¹⁸ IMO codes were removed to improve the generalizability of the algorithm, due to lack of universal availability of IMO codes in EHR warehouses.

The final list of codes and subsequent computable phenotype, that defined inclusion to the primary nephrotic syndrome cases, as well as those that defined exclusion, are listed in **Supplemental Table 1** and the algorithm scripts using SQL code logic can be accessed at <https://github.com/AoliverioUM/nephroticsyndrome-computablephenotype>. Cases of primary nephrotic syndrome were defined across all ages as subjects that were seen for at least one encounter of any kind, had 2 or more nephrotic syndrome codes, and did not have exclusion

codes for diabetes mellitus, systemic lupus erythematosus, amyloidosis, or other less common causes of secondary nephrotic syndrome such as HIV, viral hepatitis, glomerulonephritis, obstructive uropathy, or mitochondrial metabolism disorders at any time. Two or more nephrotic syndrome codes were required to minimize false positives. Additionally, for individuals younger than 20 years, codes of nephrotic syndrome not otherwise specified were also allowable for inclusion, given that the majority of children with nephrotic syndrome are steroid sensitive and consequently a kidney biopsy is pursued less frequently in this demographic. Non-cases were randomly selected individuals not meeting case criteria who were seen within the same calendar year and were within 5 years of age of the cases.

Data extraction

PCORnet® data are maintained locally with a shared structure in the Common Data Model (CDM).^{19,20} The developed computable phenotype was executed in SQL against the PCORnet® CDM at 3 academic institutions (University of Michigan, Ann Arbor, MI; The Ohio State University, Columbus, OH; Mayo Clinic, Rochester, MN, Jacksonville, FL, and Scottsdale/Phoenix, AZ) from Jan 1, 2009 to Jan 1, 2018. From the data extraction, a random selection of 50 cases and 50 non-cases was generated at each institution for chart validation.

Data coordination was conducted at Arbor Research for Health and the University of Michigan. An independent review board deemed the study exempt from IRB approval for data coordination. Institutional review board approval for this study with a waiver of individual patient consent was obtained at each individual study site where charts were abstracted.

Validation

As the gold-standard comparator, a nephrologist at each institution performed a comprehensive chart review of each generated case and non-case to determine concordance with the algorithm. Kidney biopsy data and nephrologist written documentation were carefully reviewed

for all computable phenotype-defined cases when available. Covariates collected included date of case capture by the computable phenotype, end stage kidney disease (ESKD) and transplant status at time of capture, details of kidney biopsy, and in cases of discordance between algorithm-defined and nephrologist-defined primary nephrotic syndrome, the potential causes of discordance. Following individual chart review, all three sites participated in consensus conference to facilitate any necessary additional adjudication to reach a consensus nephrologist definition.

Statistical Analyses

The classification accuracy including sensitivity, specificity, positive and negative predictive value, and F1 score of the computable phenotype was determined. Meta-analytic summary statistics for sensitivity, specificity, and area under the curve were calculated using a hierarchical summary receiver operating characteristic model.²¹ A sensitivity analysis was performed to examine the classification accuracy of the algorithm in adult patients only (20 years of age or greater). Descriptive statistics were used to describe each of the classification groups (true positives, false positives, false negatives, true negatives).

Results

At the time of execution of the computable phenotype against the PCORnet® CDM, there were 1,365,050 individual patients in the University of Michigan CDM, 1,072,626 in the Ohio State University CDM, and 1,867,416 in the Mayo Clinic CDM. The algorithm generated 666 cases of primary nephrotic syndrome at University of Michigan, 321 at Ohio State University, and 1,721 at Mayo Clinic. Classification statistics by study center are shown in **Table 1**. Of 150 cases identified as primary nephrotic syndrome by the computable phenotype, 111 were determined to be true positives by nephrologist review. Using meta-analytic summary statistics, the overall sensitivity of the computable phenotype was 99% (95% CI: 97-99%), specificity of 79% (95% CI:

74-85%), and area under the curve of 0.9 (0.84-0.97). Given that a minority of patients captured by our computable phenotype were less than 20 years old (n=37 across all 3 centers) and a computable phenotype for glomerular diseases including nephrotic syndrome in pediatric patients utilizing PEDSnet is available,¹⁷ a sensitivity analysis of the computable phenotype classification statistics solely in adults 20 years and older was performed. The performance of the computable phenotype in adults was similar and is shown in **Supplemental Table 2**.

Clinical characteristics and diagnoses for each classification category (true positive, false positive, true negative, false negative) are described in more detail in **Tables 2A/B**. The most common misclassifications were due to secondary FSGS and membranous lupus nephritis (World Health Organization Class V). Compared to true positive nephrotic syndrome cases, false positive individuals were not more likely to be on dialysis (9.9 vs. 10.2%, respectively) or have a kidney transplant (20.5 vs. 22.5%, respectively) at the time of capture by the computable phenotype than those that were correctly classified as primary nephrotic syndrome. The vast majority of both true positives and false positives were seen by nephrologists within the healthcare system queried, 99% and 97%, respectively. Amongst the 150 non-cases identified by the computable phenotype, only 8% had seen a nephrologist within the healthcare system. The sole false negative generated by the algorithm was a patient with FSGS and a kidney transplant. Of the 149 true negatives, two patients were on dialysis at the time of capture and one had received a kidney transplant.

Of the true positives, 75% (83/111) cases had biopsy reports available for review in the medical record, whereas 25% (28/111) did not. The frequency of each unique nephrotic syndrome diagnosis, by biopsy status, is shown in **Figure 1**. Membranous nephropathy was the most commonly identified diagnosis with a biopsy report available, followed by FSGS. FSGS was the most common diagnosis when a true positive primary nephrotic syndrome case was confirmed by chart validation, but a biopsy report was not available for review.

Discussion

A good computable phenotype is explicit, reproducible, reliable, and valid.²² This study demonstrates that a reliable and valid computable phenotype for primary nephrotic syndrome with an overall sensitivity of 99%, specificity of 79%, and area under the curve of 0.9 can be created using data elements from ICD-9-CM and ICD-10-CM codes alone, which are readily available across health systems in the United States. This improves the generalizability of this computable phenotype when compared to others, which have relied on IMO or SNOMED-CT codes as well. Using three academic institutions in the PCORnet® CDM, this computable phenotype captured a total of 2,708 cases of primary nephrotic syndrome; based on classification statistics determined by this study, this is estimated to represent 559 prevalent cases of true primary nephrotic syndrome at the University of Michigan, 225 at The Ohio State University, and 1,170 cases at Mayo Clinic over eight years – a significant achievement given the rarity of primary nephrotic syndrome and the fraction of the CDM that was utilized for validation.

Further refinements to the computable phenotype developed in our study may also be feasible to improve specificity. Distinguishing primary and secondary FSGS can be difficult for clinicians, and similarly proved a challenge to the computable phenotype. Secondary FSGS and focal global glomerulosclerosis collectively were a significant source of misclassification of primary nephrotic syndrome by the computable phenotype. In addition, eleven true positive cases were validated as primary FSGS despite the lack of a biopsy report available for review. While degree of foot process effacement may vary between primary and secondary maladaptive types of FSGS (e.g. obesity related FSGS), the clinical history of overt sudden onset nephrotic syndrome, response to immunosuppression, and absence of causes of secondary FSGS allowed reviewers to adequately support a diagnosis of primary FSGS.²³ Similarly, refinement of the computable phenotype for improved specificity may be possible through incorporating

medication data elements, laboratory values, or utilizing natural language processing to distinguish between primary and secondary FSGS. Natural language processing is a field of computer science and artificial intelligence that performs computational analysis of human language and is often used together with machine learning. It has emerged as a technique that can be used to identify patients with specific diagnoses by analysis of electronic medical record documentation, and has been used with success in liver²⁴ and pulmonary disease.²⁵ This approach may also be useful in distinguishing membranous nephropathy secondary to lupus nephritis. Strengthening our exclusion codes to include those for diabetes irrespective of the presence of kidney disease, may also help minimize false positives, though this may also increase false negatives in cases of true concurrent disease that may be relevant for future study (for example, when immunosuppression induces diabetes mellitus).

One limitation of our computable phenotype validation study was the inability to selectively draw non-cases from nephrology clinics at the study sites; the CDM at the time of our analysis did not contain data on specific providers or clinics within a given institution. Given that primary nephrotic syndromes are rare diseases, drawing non-cases from a general healthcare population likely inflated our sensitivity and negative predictive value statistics. For this reason, we also assessed the F1 score of the computable phenotype. An F1 score is a useful measure in classification when seeking to balance precision (the share of the predicted positive cases that are correct) and recall (the share of the actual positive cases which are predicted correctly) when there is a large number of true negatives, as expected in this population. The F1 score of the computable phenotype ranged from 0.81 to 0.9 at the individual study sites; an F1 score of 1 is considered perfect and our results indicate that the computable phenotype has an acceptable balance of precision and recall. Another limitation is that our algorithm was also developed initially within two health systems and then revised and tested specifically for PCORnet®, however there was some variation in classification statistics between study sites. This likely

represents differences in coding practices across institutions. There may also be geographical differences; while all three primary research sites are located in the Midwest, all Mayo Clinic clinical sites (Minnesota, Florida, and Arizona) contributed data to the validation study which could have lent greater variation in provider coding practices. Our study did use ICD-9-CM and ICD-10-CM codes exclusively to help with generalizability and the PCORnet® common data model to allow implementation across hundreds of hospitals, but we cannot be certain our computable phenotype will perform as well in other large data sets or EHRs. Finally, 50 cases from each site were selected at random from all ages. Given the low numbers of pediatric patients in our validation set, we are unable to differentially examine how the computable phenotype compares in adults versus children. However, a computable phenotype for pediatric nephrotic syndrome has been rigorously evaluated in PEDSnet previously and serves as another excellent resource for researchers.¹⁷

Despite these limitations, the current computable phenotype will improve upon practitioners' current abilities to identify patients with primary nephrotic syndrome for clinical trials. This will facilitate rapid and realistic assessments of site feasibility and improve enrollment for much-needed research to improve health outcomes in these rare diseases. Use of our computable phenotype and EHR data more broadly throughout PCORnet® also has the potential to strengthen pragmatic comparative effectiveness research and observational research by using "real-world" data.

In summary, we developed and validated a reliable and accurate computable phenotype for identification of primary nephrotic syndromes in children and adults using ICD-9-CM and ICD-10-CM codes. To better improve our understanding of these rare diseases, we must utilize all available tools, from genome to phenome,²⁶ and EHR data can generate large scale observational clinical data, a better understanding of how patients currently are treated and

respond to treatments, and the complications they face, augmenting our current base of knowledge.

Disclosures: L. Mariani reports the following: Consultancy Agreements: Reata Pharmaceuticals, CKD Advisory Committee; Travers Therapeutics, Advisory Board; Calliditas Therapeutics, Advisory Board; Research Funding: Boehringer Ingelheim; Honoraria: Reata Pharmaceuticals, CKD Advisory Committee; Travers Therapeutics, Advisory Board; Calliditas Therapeutics, Advisory Board; ASN BRCU; Scientific Advisor or Membership: Travers Therapeutics; Calliditas Therapeutics; Reata Pharmaceuticals. S. Almaani reports the following: Consultancy Agreements: Aurinia Pharmaceuticals; Kezar Life Sciences; Research Funding: Gilead Sciences; Scientific Advisor or Membership: Clinical Nephrology - Editorial Board; Speakers Bureau: Aurinia Pharmaceuticals. I. Ayoub reports the following: Honoraria: ACR; Scientific Advisor or Membership: Lupus Foundation of America (advisory board); Journal of Clinical Nephrology (Editorial Board). M. Denburg reports the following: Consultancy Agreements: Trisalus Life Sciences (spouse); Ownership Interest: Precision Guided Interventions LLC (spouse); In-Bore LLC (spouse); Research Funding: Mallinckrodt; Scientific Advisor or Membership: NKF Delaware Valley Medical Advisory Board; Trisalus Life Sciences Scientific Advisory Board (spouse); Other Interests/Relationships: American Society of Pediatric Nephrology Research and Program Committees; National Kidney Foundation Pediatric Education Planning Committee. H. Desmond reports the following: Research Funding: I receive a percentage of my salary from the University of Michigan but funded by the following groups: Boehringer Ingelheim. D. Gipson reports the following: Consultancy Agreements: Through University of Michigan: Vertex Pharmaceuticals; AstraZeneca; Boehringer Ingelheim; Roche/Genentech; Research Funding: Through University of Michigan: National Institutes of Health, Food and Drug Administration, Centers for Disease Control; Travers ; Novartis; Goldfinch Bio; Boehringer Ingelheim; Atrium Health Medical Foundation; Reata; Scientific Advisor or Membership: Member American Society of Nephrology; International Society of Pediatric Nephrology; American Society of Pediatric Nephrology. S. Massengill reports the following: Consultancy Agreements: Guidepoint Group; Scientific Advisor or Membership: Editorial Board of online Glomerular Disease Journal (Kargers Publishing). M. Matheny reports the following: Consultancy Agreements: NIH-VA-DoD Pain Management Grant Consortium (PMC3); Scientific Advisor or Membership: SMRB Study Section, VA HSR&D, Informatics & Methods Section; Steering Committee - Indianapolis VA HSR&D COIN Center; Steering Committee - VA HSR&D VIREC. C. Tran reports the following: Scientific Advisor or Membership: Frontiers in Pediatrics - Review Editor on the Editorial Board of Pediatric Nephrology. The remaining authors have nothing to disclose.

Funding:

- This work was supported in part by PPRN-1306-04903, PPRN Phase II, NephCure Kidney Network for Patients with Nephrotic Syndrome, Denburg (PI) 07/01/18-03/30/19.
- Andrea L. Oliverio was supported by 5T32DK007378-39 (June 2018-2019), 5KL2TR002241-04 (March 2020-Sept 2020), K23 DK123413 (Sept 2020-).
- Jonathan Troost was supported in part by the National Center for Advancing Translational Sciences (NCATS) for the Michigan Institute for Clinical and Health Research (UL1TR002240).
- Laura H. Mariani was supported by K23 DK115891.

Author Contributions: Andrea Oliverio: Data curation; Formal analysis; Project administration; Validation; Writing - original draft; Writing - review and editing. Dorota Marchel: Data curation; Validation; Writing - review and editing. Jonathon Troost: Formal analysis; Methodology; Writing - review and editing. Isabelle Ayoub: Data curation; Validation; Writing - review and editing. Salem Almaani: Data curation; Validation; Writing - review and editing. Jessica Greco: Data curation; Validation; Writing - review and editing. Cheryl Tran: Data curation; Validation; Writing - review and editing. Michelle Denburg: Conceptualization; Funding acquisition; Methodology; Writing - review and editing. Michael Matheny: Conceptualization; Methodology; Software; Writing - review and editing. Chad Dorn: Methodology; Software; Writing - review and editing. Susan Massengill: Data curation; Writing - review and editing. Hailey Desmond: Data curation; Project administration; Validation; Writing - review and editing. Debbie Gipson: Conceptualization; Data curation; Formal analysis; Methodology; Supervision; Writing - review and editing. Laura Mariani: Conceptualization; Data curation; Formal analysis; Funding acquisition; Resources; Supervision; Validation; Visualization; Writing - review and editing.

References

1. Group KDIGOKGW. KDIGO Clinical Practice Guideline for Glomerulonephritis. *Kidney inter, Suppl* 2012; **2**: 139-274.
2. Friedman DJ, Pollak MR. APOL1 Nephropathy: From Genetics to Clinical Applications. *Clin J Am Soc Nephrol* 2020.
3. Merchant ML, Barati MT, Caster DJ, et al. Proteomic Analysis Identifies Distinct Glomerular Extracellular Matrix in Collapsing Focal Segmental Glomerulosclerosis. *J Am Soc Nephrol* 2020; **31**(8): 1883-904.
4. Mariani LH, Martini S, Barisoni L, et al. Interstitial fibrosis scored on whole-slide digital imaging of kidney biopsies is a predictor of outcome in proteinuric glomerulopathies. *Nephrol Dial Transplant* 2018; **33**(2): 310-8.
5. Troost JP, Trachtman H, Spino C, et al. Proteinuria Reduction and Kidney Survival in Focal Segmental Glomerulosclerosis. *Am J Kidney Dis* 2021; **77**(2): 216-25.
6. Laurin LP, Gasim AM, Derebail VK, et al. Renal Survival in Patients with Collapsing Compared with Not Otherwise Specified FSGS. *Clin J Am Soc Nephrol* 2016; **11**(10): 1752-9.
7. McGrogan A, Franssen CF, de Vries CS. The incidence of primary glomerulonephritis worldwide: a systematic review of the literature. *Nephrol Dial Transplant* 2011; **26**(2): 414-30.
8. Carter SA, Gutman T, Logeman C, et al. Identifying Outcomes Important to Patients with Glomerular Disease and Their Caregivers. *Clin J Am Soc Nephrol* 2020; **15**(5): 673-84.
9. Collaboratory NHCSR. Electronic Health Records-Based Phenotyping. <https://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/> (accessed November 11, 2020 2020).
10. Koola JD, Davis SE, Al-Nimri O, et al. Development of an automated phenotyping algorithm for hepatorenal syndrome. *J Biomed Inform* 2018; **80**: 87-95.
11. Forrest CB, McTigue KM, Hernandez AF, et al. PCORnet(R) 2020: current state, accomplishments, and future directions. *J Clin Epidemiol* 2021; **129**: 60-7.
12. Wiese AD, Roumie CL, Buse JB, et al. Performance of a computable phenotype for identification of patients with diabetes within PCORnet: The Patient-Centered Clinical Research Network. *Pharmacoepidemiol Drug Saf* 2019; **28**(5): 632-9.
13. Bachmann KN, Roumie CL, Wiese AD, et al. Diabetes medication regimens and patient clinical characteristics in the national patient-centered clinical research network, PCORnet. *Pharmacol Res Perspect* 2020; **8**(5): e00637.

14. Tison GH, Chamberlain AM, Pletcher MJ, et al. Identifying heart failure using EMR-based algorithms. *Int J Med Inform* 2018; **120**: 1-7.
15. Services CfMaM. Chronic conditions data warehouse. <https://www2.ccwdata.org/web/guest/condition-categoriesFebruary> 18, 2021).
16. HCUP CCS March 2017. <https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp> (accessed November 12, 2020).
17. Denburg MR, Razzaghi H, Bailey LC, et al. Using Electronic Health Record Data to Rapidly Identify Children with Glomerular Disease for Clinical Research. *J Am Soc Nephrol* 2019; **30**(12): 2427-35.
18. Impact of ICD-10-CM/PCS on Research Using Administrative Databases 2016. <https://www.hcup-us.ahrq.gov/reports/methods/methods.jsp> (accessed April 2, 2021).
19. Collins FS, Hudson KL, Briggs JP, Lauer MS. PCORnet: turning a dream into reality. *J Am Med Inform Assoc* 2014; **21**(4): 576-7.
20. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc* 2014; **21**(4): 578-82.
21. Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *J Clin Epidemiol* 2004; **57**(9): 925-32.
22. Richesson RSM. Electronic Health Records-Based Phenotyping <https://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/#intro-definitions> (accessed November 12, 2020).
23. De Vriese AS, Sethi S, Nath KA, Glassock RJ, Fervenza FC. Differentiating Primary, Genetic, and Secondary FSGS in Adults: A Clinicopathologic Approach. *J Am Soc Nephrol* 2018; **29**(3): 759-74.
24. Chang EK, Yu CY, Clarke R, et al. Defining a Patient Population With Cirrhosis: An Automated Algorithm With Natural Language Processing. *J Clin Gastroenterol* 2016; **50**(10): 889-94.
25. Weiner M, Dexter PR, Heithoff K, et al. Identifying and Characterizing a Chronic Cough Cohort Through Electronic Health Records. *Chest* 2020.
26. Mariani LH, Kretzler M. Pro: 'The usefulness of biomarkers in glomerular diseases'. The problem: moving from syndrome to mechanism--individual patient variability in disease presentation, course and response to therapy. *Nephrol Dial Transplant* 2015; **30**(6): 892-8.

Table 1. Performance characteristics of the computable phenotype at 3 academic PCORnet centers

<i>Center</i>	True positive	False positive	False negative	True negative	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Accuracy	F1 score	Area under curve (95% CI)
<i>University of Michigan</i>	42	8	1	49	0.98	0.86	0.84	0.98	0.91	0.90	0.92 (0.87- 0.97)
<i>The Ohio State University</i>	35	15	0	50	1.00	0.77	0.70	1.00	0.85	0.82	0.88 (0.83- 0.94)
<i>Mayo Clinic</i>	34	16	0	50	1.00	0.76	0.68	1.00	0.84	0.81	0.88 (0.83- 0.93)

Table 2A. Diagnosis and demographic characteristics of true and false positives identified by the computable phenotype

	True Positives N=111	False Positives N=39
Nephrologist Validated Diagnosis n (%)		
<i>MCD</i>	18 (16.2%)	0
<i>FSGS</i>	38 (34.2%)	0
<i>MN</i>	45 (40.5%)	0
<i>SSNS/SRNS</i>	4 (3.6%)	0
<i>Other^a</i>	6 (5.4%)	0
<i>Secondary FSGS</i>	0	9 (23.1%)
<i>Lupus nephritis</i>	0	9 (23.1%)
<i>Diabetic kidney disease</i>	0	4 (10.3%)
<i>Focal global glomerulosclerosis</i>	0	2 (5.1%)
<i>Vasculitis</i>	0	2 (5.1%)
<i>Hypertensive nephrosclerosis</i>	0	1 (2.6%)
<i>Other secondary cause^b</i>	0	9 (23.1%)
<i>Unknown</i>	0	3 (7.7%)
ESKD status n (%)		
<i>Dialysis</i>	11 (9.9%)	4 (10.26%)
<i>Kidney Transplant</i>	25 (22.5%)	8 (20.5%)
Age yrs, mean (SD)	43.3 (20.9)	44.9 (19.2)
Nephrology Encounter at Site n (%)	110 (99%)	38 (97.4%)
Time from Nephrology Encounter to Computable Phenotype Capture	396 (0-1792)	543 (16–2279)
Days, median (IQR)		
Biopsy status n (%)		N/A
<i>Documented and Reviewed</i>	83 (75%)	
<i>Not documented or Not performed</i>	28 (25%)	

MCD: minimal change disease. FSGS: focal segmental glomerulosclerosis. MN: membranous nephropathy. SSNS/SRNS: steroid sensitive nephrotic syndrome, steroid resistant nephrotic syndrome. ESKD: end stage kidney disease. ^a Other primary etiologies included four cases of congenital nephrotic syndrome, one membranoproliferative glomerulonephritis, and one lipoprotein glomerulonephropathy. ^b Other secondary etiologies included two cases of interstitial nephritis, and one each of secondary nephrotic syndrome due to lymphoma, Alport syndrome, obstructive nephropathy, acute mediated rejection in transplant, immune complex glomerulonephritis, hemolytic uremic syndrome, and Kawasaki's disease.

Table 2B. Diagnosis and demographic characteristics of true and false negatives identified by the computable phenotype

	True Negatives N=149	False Negatives N=1
Nephrologist Validated		
Diagnosis n (%)		
<i>MCD</i>	0	0
<i>FSGS</i>	0	1
<i>MN</i>	0	0
<i>SSNS/SRNS</i>	0	0
ESKD status		
<i>Dialysis</i>	2	0
<i>Kidney Transplant</i>	1	1
Age yrs, mean (SD)	44.3 (20.3)	29
Nephrology Encounter at Site n (%)	11 (7.4%)	1 (100%)

MCD: minimal change disease. FSGS: focal segmental glomerulosclerosis. MN: membranous nephropathy. SSNS/SRNS: steroid sensitive nephrotic syndrome, steroid resistant nephrotic syndrome. ESKD: end stage kidney disease.

Figure 1. Frequency of true positive primary nephrotic syndrome diagnoses, by biopsy review status

A. n=83 true positive primary nephrotic syndrome cases with biopsies reviewed by chart validation. Other diagnoses consisted of one of each: congenital nephrotic syndrome, lipoprotein glomerulopathy, and membranoproliferative glomerulonephritis.

B. n=28 true positive primary nephrotic syndrome cases where biopsies were not performed or unable to be reviewed by chart validation. Other diagnoses consisted of 3 cases of congenital nephrotic syndrome.

MCD: minimal change disease; FSGS: focal segmental glomerulosclerosis; MN: membranous nephropathy; SSNS: steroid sensitive nephrotic syndrome; SRNS: steroid resistant nephrotic syndrome

Figure 1A

Figure 1A.

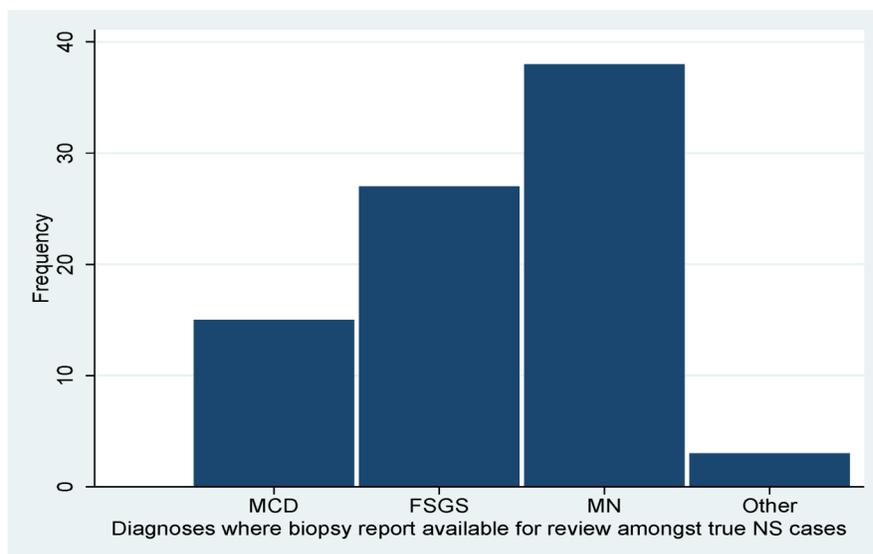


Figure 1B.

