

ARTICLE

Automated data extraction model for the USIDNET registry: Bigger, faster, and better data collection

Vaibhavi Vichare¹, Gonench Kilich², Charlotte Cunningham-Rundles³, Michelle Onyekaba⁴, Roshini S. Abraham⁵, Jennifer Puck⁶, Ramsay Fuleihan⁷, Rebecca A. Marsh^{8,9}, and Kathleen E. Sullivan³

The United States Immunodeficiency Network (USIDNET) is an NIH-funded research consortium that advances scientific investigation on inborn errors of immunity (IEIs). Formerly, USIDNET registry data were collected via an opt-in system with patient informed consent. The current registry uses semiautomated, de-identified data extraction from EPIC with a consent waiver. This study was performed to assess whether the new method improved enrollment using data from one site. Diagnoses, sex, and age within the new ($n = 2,796$) and old ($n = 551$) registries were defined. The new registry enrolled more subjects and had six times more clinical features recorded per patient on average and 22 times more laboratory data recorded per patient. Response to queries is much more rapid, with execution of a database query within a day. There were differences in enrollment demographics depending on underlying diagnosis. The design of the new USIDNET registry may better capture a greater number and representation of patients compared to the old registry.

Introduction

The United States Immunodeficiency Network (USIDNET) was designed to provide multiple resources to the clinical immunology community. A registry is a central component of the suite of resources. The previous version of the USIDNET registry comprised over 5,000 inborn error of immunity (IEI) patients from 43 centers (1). That registry collected clinical, genetic, and phenotypic patient data via a clinician accessed REDCap data entry system. USIDNET registry data have been used as a scientific resource for critical research on the molecular and genetic pathways involved in IEI biology (selected examples [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]); however, longitudinal data were unstructured and incompletely provided. Antibody deficiencies represented about half of all patients in the previous registry. A goal of USIDNET has been to improve longitudinal data collection, data completeness, and patient representation. Longitudinal data collection will improve outcomes research, and increasing enrollment and types of patients is critical to define IEI phenotypes among different populations. A 2024 study by O'Connell and colleagues compared a cohort of pediatric patients diagnosed with IEI to a comparator congenital heart disease (CHD) cohort and found that while the CHD cohort demographics mirrored the local

population, the IEI cohort disproportionately comprised European ancestry patients with private insurance. Specifically, they found varying diagnostic rates of IEI between different demographics and differences associated with emergency room visits and hospitalizations (15). Another study found deaths related to IEI had the highest rate for non-European ancestry patients (16). Challenges related to obtaining a referral to a tertiary care specialist to receive an IEI diagnosis may contribute to some differences in outcomes among health disparity populations (17). Recognizing these challenges, we set out to try to improve data collection such that USIDNET would improve enrollment.

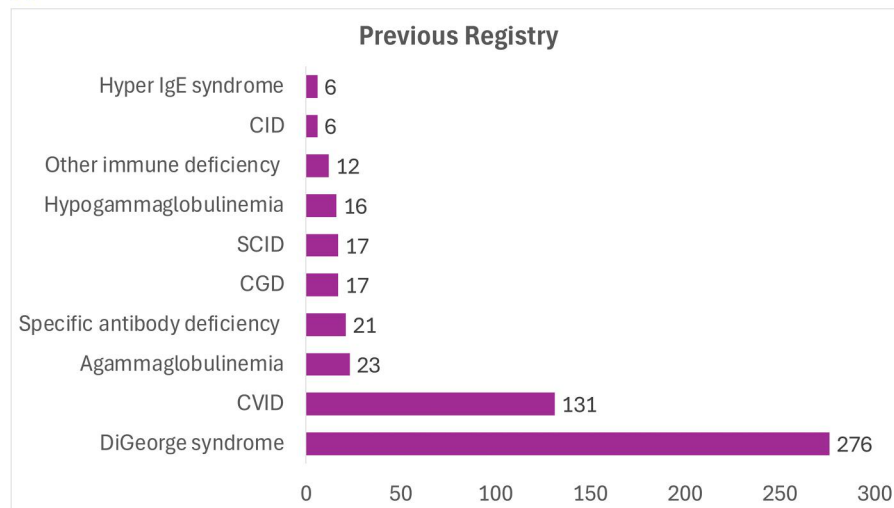
The new USIDNET registry uses automated data extraction to improve longitudinal data and a waiver of consent to improve representation of the clinic population at participating sites. (Fig. S1) This study was performed at a single site since the current participating sites do not match the sites participating in the previous registry. We compared diagnoses and enrollment and then compared them to inpatient and outpatient populations, as well as comparing the site's demographics in the previous and current registries.

¹Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; ²Division of Allergy Immunology, Children's Hospital of Philadelphia, Philadelphia, PA, USA; ³Department of Medicine, Department of Pediatrics, Icahn School of Medicine at Mount Sinai New York, NY, USA; ⁴University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA; ⁵Department of Pathology and Laboratory Medicine, Nationwide Children's Hospital, Columbus, OH, USA; ⁶Division of Allergy/Immunology and Blood and Marrow Transplantation, Department of Pediatrics, UCSF Benioff Children's Hospital, University of California San Francisco, San Francisco, CA, USA; ⁷Division of Allergy, Immunology and Rheumatology, Department of Pediatrics, Columbia University Irving Medical Center, New York, NY, USA; ⁸Division of Bone Marrow Transplantation and Immune Deficiency, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; ⁹Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA.

Correspondence to Kathleen E. Sullivan: sullivanke@chop.edu.

© 2025 Vichare et al. This article is available under a Creative Commons License (Attribution 4.0 International, as described at <https://creativecommons.org/licenses/by/4.0/>).

A



B

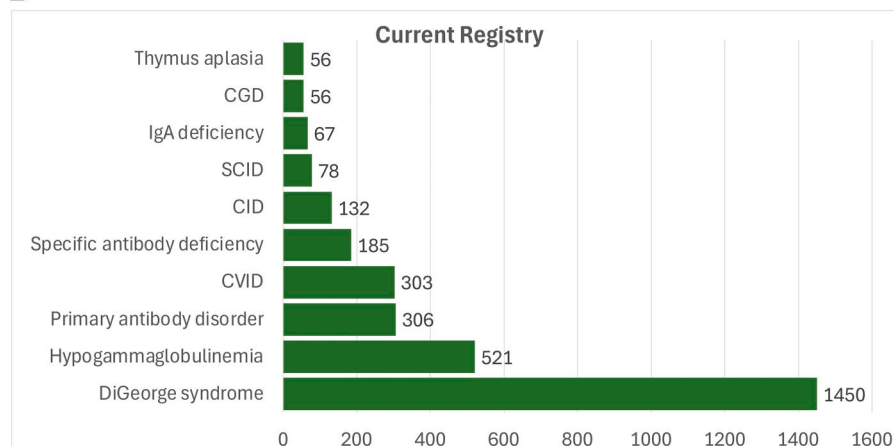


Figure 1. The 10 most common diagnoses in each registry. These two graphs list the top 10 diagnoses as unique patient counts. **(A)** The most common diagnoses in the previous registry were DiGeorge syndrome ($n = 276$), CVID ($n = 131$), agammaglobulinemia ($n = 23$), SCID ($n = 17$), and CGD ($n = 17$). **(B)** The most common diagnoses in the current registry ($n = 2,796$) were DiGeorge syndrome ($N = 1,450$), hypogammaglobulinemia ($N = 521$), primary antibody disorder ($N = 306$), CVID ($N = 303$), and specific antibody deficiency ($N = 185$).

Results

In this study, we hypothesized that implementing a semi-automated data extraction method from the electronic medical record with a waiver of consent would improve enrollment. This protocol was deemed “not Human Subject Research” by Children’s Hospital of Philadelphia because the extracted data do not include personal health information. The new registry roughly doubled enrollment at this one site from 551 in 2019 to 2,796 enrollees in 2025. This represents an 82% capture of current patients identified as immunodeficient at this institution. We assessed disease categories by physician definition in the previous USIDNET registry and International Classification of Diseases (ICD) code definition in the current registry (Figs. 1 and S1). The distribution was different. Previously, IEI types were restricted to the pull down list of options. USIDNET V2 allows any disorder defined by an ICD10 code. The density of data per patient also increased. As an example, the previous registry had 131 unique patients with common variable immunodeficiency (CVID). The average number of non-laboratory data points per patient was 13 compared to 59 per patient in the new registry, which had 303 patients with CVID. Overall, the patients in the previous registry had 6.5 clinical features recorded per patient

and 26 lab results per patient. In the new registry, the patients had 38.5 clinical features and 866 lab values on average ($P < 0.0001$ for both). This represents a sixfold increase in clinical features recorded per patient and a 22-fold increase in laboratory data.

The structure of the previous registry required significant manual data curation of the output. The data queries can be completed in less than a day from the new registry. We tested the fidelity of the diagnosis codes in our new registry. Six of the people with a CVID diagnosis also had a diagnosis of IgA deficiency.

We additionally hypothesized that the demographics of the enrolled patients would change under the new model. The median age in the previous registry was 28 years of age (minimum 9 years, maximum 72 years of age) at time the registry closed to enrollment (2020). The median age in the current registry was 17 years of age (minimum 0 years, maximum 54 years of age) (data not shown). The previous registry was 63.7% male, and the current registry was 61.6% male. We found that there was a statistically significant ($P < 0.0001$) change in the distribution of different populations in the new cohort as compared to the old, with an increase in the proportion of people of color with the

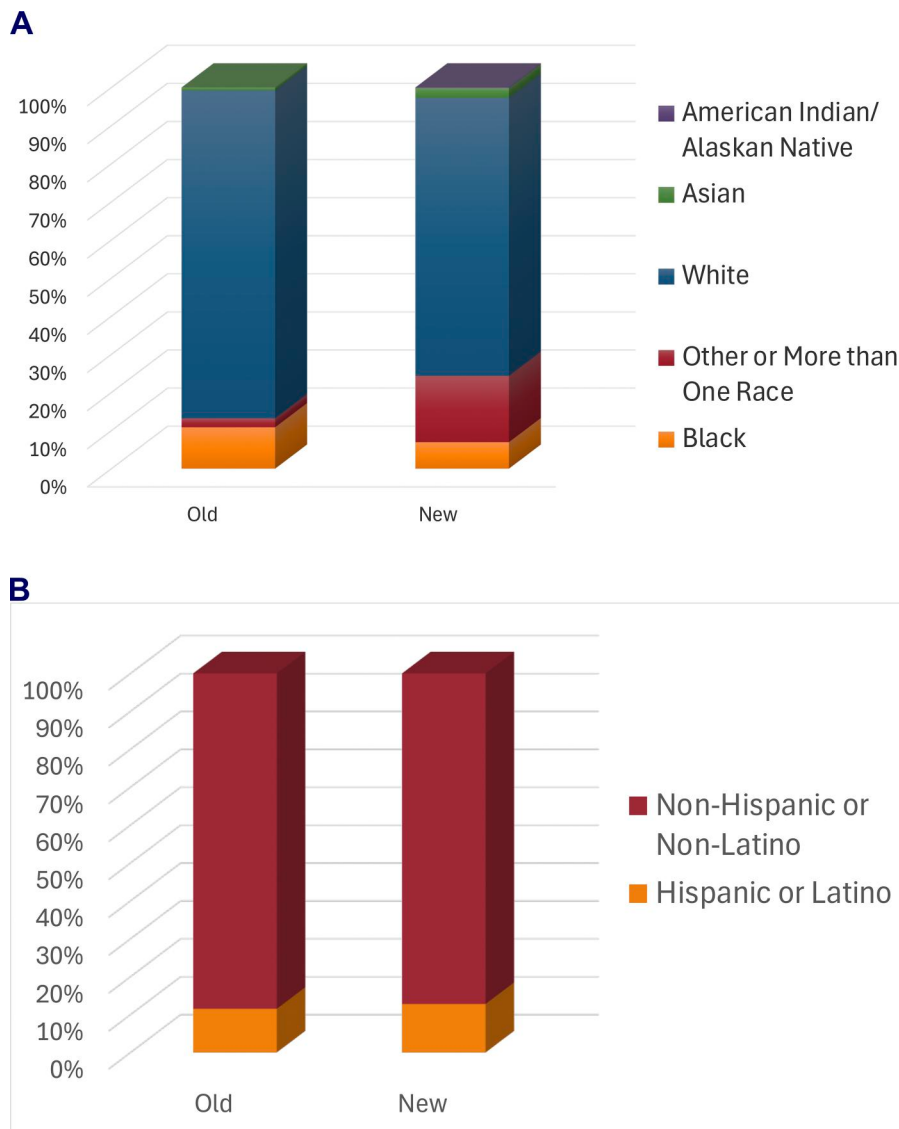


Figure 2. Demographics in previous and current registries. (A) The current (new) registry data from the study institution includes racial information: 0.18% American Indian/Alaska Native, 2.55% Asian, 6.94% Black, 72.93% white, and 17.40% other or more than one race. The previous (old) registry has the following racial information: 0.78% Asian, 10.85% Black, 86.05% White, and 2.33% other or more than one race (Chi-square $P < 0.0001$). **(B)** The current (new) registry includes ethnicity information: 12.78% Hispanic/Latino, 87.22% non-Hispanic/Latino. The previous (old) registry included ethnicity information on: 11.50% Hispanic/Latino, 88.50% non-Hispanic/Latino (Chi-square $P = 0.697$).

largest increase in people identifying as “more than one race” (Fig. 2 A). The population of people of Latin ancestry changed very little between the two registries (Fig. 2 B).

To understand if there were disease-specific effects, we assessed the racial and ethnic representation in agammaglobulinemia, severe combined immunodeficiency (SCID), and chronic granulomatous disease (CGD) in both cohorts. In the current registry, all patients with SCID would have been detected through newborn screening, and therefore ascertainment should reflect the population of the region. Agammaglobulinemia and CGD both require recognition and referral to a medical center, a process where challenges might intervene and lead to imperfect ascertainment. We found there was a statistically significant difference in distribution according to ancestry between the two registries for agammaglobulinemia ($P = 0.0194$), but not for SCID ($P = 0.43$) or CGD ($P = 0.62$). The distribution in the three disease states was similar to the overall cohort distribution in the recent registry with higher numbers of “more than one race” in all three disease categories than in the previous

registry. Latin heritage and sex distribution remained relatively unchanged (not shown).

To better understand the demographics of the broader population being treated within this institution, we looked at the breakdown of total visits between 2010 and 2024 at this one institution. We found that there were fewer patients of color seen in the immunology, gastroenterology, and oncology subspecialties relative to neonatology and the emergency department (Fig. 3). The current registry demographics differed from the total population ($P < 0.0001$) but not from the immunology clinic demographics. The total immunology clinic visits included: 16.49% Black, 14.77% other, 62.86% white, 3.36% Asian, 0.16% American Indian/Alaskan Native, and 1.46% unknown.

Discussion

We hypothesized that automated data extraction and a waiver of consent would remove some challenges to participation and improve the range of patients included in the USIDNET registry.

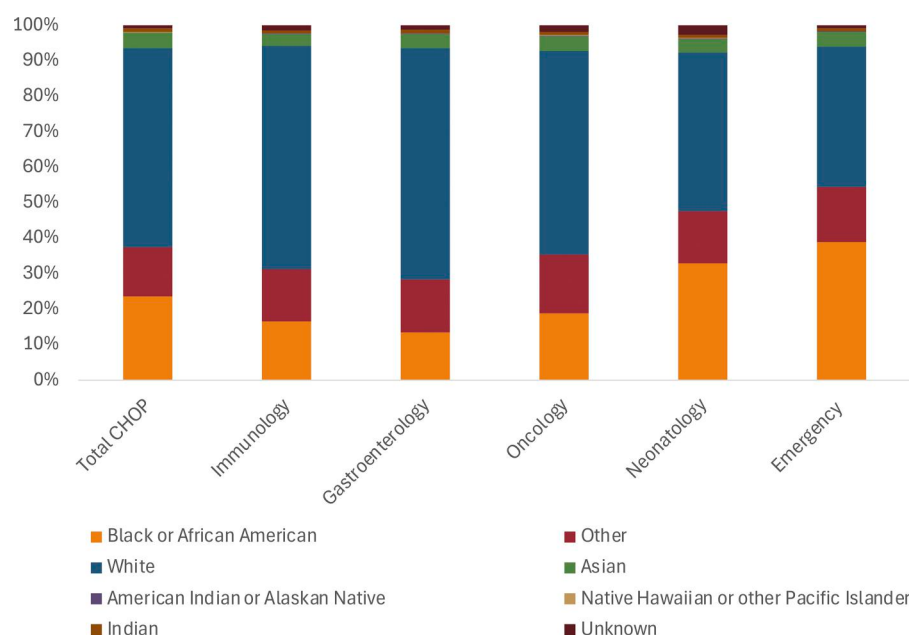


Figure 3. Comparison of demographics across different clinical settings. CHOP visits were tabulated by department between 2010 and 2024 using Slicer Dicer across different clinical groups. The demographic distribution varied across clinical sites, with neonatology and the emergency department having the highest rates of non-European ancestry patients. The difference between total CHOP and immunology was significantly different with $P < 0.0001$, whereas the difference between immunology and the current registry was not significant.

Clear improvements in expanding the range of patients were observed. The number of patients from this single institution doubled between the previous and current registries, and the current registry collected 82% of all patients identified through ICD codes as having an immunodeficiency. It is not possible to define the variables governing this increase, but removing the burden of data entry and consenting likely played a significant role.

We noted a more granular listing of antibody disorders, which likely reflects a methodologic change in data collection and suggests that USIDNET successfully expanded the range of diagnoses in the current version compared to the previous version. Advances in artificial intelligence may allow better identification of patients and better assignment to discrete diagnostic groups (18, 19, 20).

The percentage of patients from non-European ancestry backgrounds also changed in the current registry. It is useful to compare the demographics to other large populations in the U.S. In a study of national insurance claims, 83% of individuals were European ancestry (21). In a Chicago study using ICD codes, 63% of individuals with IEI were European ancestry (22), and a New York study using ICD codes found that 77% were European ancestry (23). The previous USIDNET registry had 86% European ancestry participants with the current registry having 73%.

The current registry collects data back to 2018, a time frame when newborn screening for SCID was occurring in the state of Pennsylvania. Therefore, comprehensive capture of all patients in our catchment area would be expected. In the previous registry, 13% non-European ancestry patients with SCID were recorded compared to 42% in the current registry. This is similar to the overall clinic population of 37% non-European ancestry patients but somewhat less than the overall Children's Hospital of Philadelphia (CHOP) population with 47% non-European ancestry patients. For agammaglobulinemia, the previous registry had 18% non-European ancestry patients compared to 57% in the

current registry. Thus, progress has been made in improving USIDNET enrollment at one institution.

These data demonstrate that the current registry enrollment demographics are similar to the subspecialty clinics at this institution. However, we note a substantial difference between the three outpatient specialties analyzed compared to the total institutional population, the inpatient neonatology population, and the emergency department population. There are likely multifactorial reasons for the USIDNET enrollees not matching the overall institutional demographics. Health disparity populations have less access to tertiary care, likely leading to fewer diagnoses of immunodeficiencies. We noted that gastroenterology and oncology had similar demographics, supporting an element of diminished access in general to subspecialists for people of non-European ancestry. Ultimately, the new system of USIDNET data extraction led to a statistically significant increase in the enrollment of patients and improved data density.

This study examined an important question for all registries. Are the data representative of what clinicians will encounter in their practice, and are they therefore of value? USIDNET has made major strides in improving number of patients and demographic matching to the clinical anchor population, at least at one institution. Furthermore, data extraction is simpler, and the longitudinal data are structured. This study does have some limitations, however. The enrollment in the previous registry was smaller than the current registry, leading to small numbers of evaluable patients for some analytic categories. ICD10 codes represent the manner in which a patient is categorized as having an IEI. All patients are under care in an immunology clinic, thereby providing confidence in correct coding; however, ICD10 codes are incomplete for IEI. Nevertheless, these data support our hypothesis that representation will improve if challenges and hurdles are minimized for inclusion. There remains significant work to provide a fully representational data set, but the recent progress demonstrates that USIDNET provides data on a population of value to practicing clinicians.

Materials and methods

Patient enrollment into USIDNET comes from immunology clinics and an institutional immunologist. USIDNET leverages the Translational Data Warehouse built on PostgreSQL. This infrastructure supports automated data extraction through standardized query templates applied to EPIC Clarity, allowing consistent and reliable integration of clinical data from participating sites. It is a standardize database with R clinical concepts and captures more than 50 data fields, including lab data, imaging, pathology, medications, procedures, and visit diagnoses. Demographic data are extracted directly from Clarity. The Clarity data warehouse includes all orders, codes required for billing compliance, such as ICD code for the visit, ICD codes for all laboratory and medication orders, and all results appearing as discrete fields in EPIC. Demographic data (including death and insurance type) and insurance are also captured. ICD codes are hardwired into EPIC and occur as a pull-down list for the most common billing codes in each specialty clinic, with the capacity to free-type in a condition, and the system generates a list of conditions and ICD codes, which are refined with additional typing, or the ICD code may be selected from the list by clicking it. Diagnoses are defined by ICD10 code.

We analyzed data from a single site, Children's Hospital of Philadelphia, that had participated in both registries and for which we could obtain demographic data across different clinical settings. The data were de-identified, and therefore this study received an exemption from the Institutional Review Board. For the previous USIDNET Children's Hospital of Philadelphia Registry ($n = 551$), patient demographic data were obtained following a consent process and data entry by a trained research coordinator or a clinician. For the new registry ($n = 2,796$), patient data were collected from the electronic health record using a semiautomated data extraction model with a consent waiver. Patients without data for a given demographic were excluded from the analysis for that demographic. The data were analyzed based on an extraction from September 30, 2025. Demographic data on total CHOP patient visits and visits to immunology, gastroenterology, oncology, neonatology, and emergency medicine between 2010 and 2024 were ascertained using the SlicerDicer Epic tool. Chi-squared and Fisher's exact test were completed using GraphPad Prism to determine statistical significance in demographic variance between the old and new registry as well between total CHOP and CHOP immunology visits.

Data availability

All USIDNET data are freely available upon approval by the Steering Committee. A standardized query form is on the USIDNET website.

Acknowledgments

The authors acknowledge all the patients who have taught us over the years.

Author contributions: Vaibhavi Vichare: data curation and software. Gonench Kilich: investigation, project administration, and writing—review and editing. Charlotte Cunningham-

Rundles: data curation, funding acquisition, investigation, methodology, resources, and writing—review and editing. Michelle Onyekaba: data curation and formal analysis. Roshini S. Abraham: data curation, resources, supervision, and writing—review and editing. Jennifer Puck: conceptualization, investigation, supervision, and writing—review and editing. Ramsay Fuleihan: methodology and writing—review and editing. Rebecca A. Marsh: conceptualization, project administration, and writing—review and editing. Kathleen E. Sullivan: conceptualization, data curation, investigation, methodology, project administration, resources, supervision, and writing—review and editing.

Disclosures: R.S. Abraham reported grants from Amgen and personal fees from Novartis outside the submitted work. J. Puck reported "other" from UpToDate outside the submitted work. R. Fuleihan reported personal fees from Takeda and Grifols outside the submitted work. R.A. Marsh reported "other" from Pharming Healthcare outside the submitted work, and is employed part-time by Pharming Healthcare, Warren, NJ. K.E. Sullivan reported personal fees from Immune Deficiency Foundation outside the submitted work. No other disclosures were reported.

Submitted: 12 June 2025

Revised: 3 October 2025

Accepted: 7 November 2025

References

1. Sullivan, K.E., J.M. Puck, L.D. Notarangelo, R. Fuleihan, T. Caulder, C. Wang, M. Boyle, and C. Cunningham-Rundles. 2014. USIDNET: A strategy to build a community of clinical immunologists. *J. Clin. Immunol.* 34: 428–435. <https://doi.org/10.1007/s10875-014-0028-1>
2. Yonkof, J., A.O. Gupta, P. Fu, E. Garabedian, and J.D. Dalal. 2017. Early transplant improves outcomes in patients with chronic granulomatous disease: Results from United States immunodeficiency Network (USIDNET). *Blood.* 130:671. https://doi.org/10.1182/blood.v130.suppl_1.671.671
3. Tuano, K.S., J.S. Orange, K. Sullivan, C. Cunningham-Rundles, F.A. Bonilla, and C.M. Davis. 2015. Food allergy in patients with primary immunodeficiency diseases: Prevalence within the US immunodeficiency Network (USIDNET). *J. Allergy Clin. Immunol.* 135:273–275. <https://doi.org/10.1016/j.jaci.2014.09.024>
4. Gernez, Y., A.F. Freeman, S.M. Holland, E. Garabedian, N.C. Patel, J.M. Puck, K.E. Sullivan, J. Akhter, E. Secord, K. Chen, et al. 2018. Autosomal dominant hyper-IgE Syndrome in the USIDNET registry. *J. Allergy Clin. Immunol. Pract.* 6:996–1001. <https://doi.org/10.1016/j.jaip.2017.06.041>
5. O'Toole, D., D. Groth, H. Wright, F.A. Bonilla, R.L. Fuleihan, C. Cunningham-Rundles, K.E. Sullivan, H.D. Ochs, R. Marsh, and E. Feuille. 2022. X-linked agammaglobulinemia: Infection frequency and infection-related mortality in the USIDNET registry. *J. Clin. Immunol.* 42:827–836. <https://doi.org/10.1007/s10875-022-01237-1>
6. Patrawala, M., Y. Cui, L. Peng, R.L. Fuleihan, E.K. Garabedian, K. Patel, and L. Gugliani. 2020. Pulmonary disease burden in primary immune deficiency disorders: Data from USIDNET registry. *J. Clin. Immunol.* 40: 340–349. <https://doi.org/10.1007/s10875-019-00738-w>
7. Ruffner, M.A., USIDNET Body Weight Group, and K.E. Sullivan. 2018. Complications associated with underweight primary immunodeficiency patients: Prevalence and associations within the USIDNET registry. *J. Clin. Immunol.* 38:283–293. <https://doi.org/10.1007/s10875-018-0492-0>
8. Rohr, K., and J. Yu. 2018. Impact of pulmonary complications on quality of life in the USIDNET registry. *J. Clin. Immunol.* 38:330–444.
9. Méndez Barrera, J.A., S. Rocha Guzman, E. Hierro Cascajares, E.K. Garabedian, R.L. Fuleihan, K.E. Sullivan, S.O. Lugo Reyes. 2023. Who's your data? Primary immune deficiency differential diagnosis prediction via machine learning and data mining of the USIDNET registry. *Clin. Immunol.* 255:109759. <https://doi.org/10.1016/j.clim.2023.109759>

10. Padem, N., H. Wright, R. Fuleihan, E. Garabedian, D. Suez, C. Cunningham-Rundles, R.A. Marsh, and A. Khojah. 2022. Rheumatologic diseases in patients with inborn errors of immunity in the USIDNET registry. *Clin. Rheumatol.* 41:2197–2203. <https://doi.org/10.1007/s10067-021-06044-4>
11. Smith, K.L., D. Dai, B.P. Modi, R. Sara, E. Garabedian, R.A. Marsh, J. Puck, E. Secord, K.E. Sullivan, S.E. Turvey, et al. 2022. Inborn errors of immunity associated with type 2 inflammation in the USIDNET registry. *Front. Immunol.* 13:831279. <https://doi.org/10.3389/fimmu.2022.831279>
12. Durkee-Shock, J., A. Zhang, H. Liang, H. Wright, J. Magnusson, E. Garabedian, R.A. Marsh, K.E. Sullivan, M.D. Keller, and USIDNET Consortium. 2022. Morbidity, mortality, and therapeutics in combined immunodeficiency: Data from the USIDNET registry. *J. Allergy Clin. Immunol. Pract.* 10:1334–1341.e6. <https://doi.org/10.1016/j.jaip.2022.01.042>
13. Deshpande, D.R., Y.Y. Demirdag, R.A. Marsh, K.E. Sullivan, J.S. Orange, and USIDNET Consortium. 2021. Relationship between severity of T cell lymphopenia and immune dysregulation in patients with DiGeorge Syndrome (22q11.2 deletions and/or related TBX1 mutations): A USIDNET study. *J. Clin. Immunol.* 41:29–37. <https://doi.org/10.1007/s10875-020-00854-y>
14. McDonnell, J.C., K. Cousins, M.E.M. Younger, A. Lane, H. Abolhassani, R.S. Abraham, S. Al-Tamemi, J.C. Aldave-Becerra, E.H. Al-Faris, A. Alfaro-Murillo, et al. 2023. COVID-19 vaccination in patients with inborn errors of immunity reduces hospitalization and critical care needs related to COVID-19: A USIDNET report. *J. Clin. Immunol.* 44:86. <https://doi.org/10.1007/s10875-023-01613-5>
15. O'Connell, P., O. Stewart, and D. Bogunovic. 2024. Racial disparities in the diagnosis of inborn errors of immunity. *J. Clin. Immunol.* 44:186. <https://doi.org/10.1007/s10875-024-01790-x>
16. Ong, M.S., N.L. Rider, S. Stein, P.J. Maglione, A. Galbraith, D.V. DiGiuseppe, and J.R. Farmer. 2024. Racial and ethnic disparities in early mortality among patients with inborn errors of immunity. *J. Allergy Clin. Immunol.* 153:335–340.e1. <https://doi.org/10.1016/j.jaci.2023.09.036>
17. Wallace, L.J., M.S. Ware, C. Cunningham-Rundles, R.L. Fuleihan, and P.J. Maglione. 2021. Clinical disparity of primary antibody deficiency patients at a safety net hospital. *J. Allergy Clin. Immunol. Pract.* 9:2923–2925.e1. <https://doi.org/10.1016/j.jaip.2021.03.021>
18. Rider, N.L., M. Coffey, A. Kurian, J. Quinn, J.S. Orange, V. Modell, and F. Modell. 2023. A validated artificial intelligence-based pipeline for population-wide primary immunodeficiency screening. *J. Allergy Clin. Immunol.* 151:272–279. <https://doi.org/10.1016/j.jaci.2022.10.005>
19. Johnson, R., A.V. Stephens, R. Mester, S. Knyazev, L.A. Kohn, M.K. Freund, L. Bondhus, B.L. Hill, T. Schwarz, N. Zaitlen, et al. 2024. Electronic health record signatures identify undiagnosed patients with common variable immunodeficiency disease. *Sci. Transl. Med.* 16:eade4510. <https://doi.org/10.1126/scitranslmed.ade4510>
20. Riviere, J.G., R. Cantenys-Saba, G. Carot-Sans, J. Piera-Jimenez, M.J. Butte, P. Soler-Palacin, X.P. Peng. 2025. Current perspectives and challenges of using artificial intelligence in immunodeficiencies. *J. Allergy Clin. Immunol.* 156:878–888. <https://doi.org/10.1016/j.jaci.2025.06.015>
21. Rider, N.L., A. Truxton, T. Ohrt, I. Margolin-Katz, M. Horan, H. Shin, R. Davila, V. Tenenbaum, J. Quinn, V. Modell, et al. 2024. Validating inborn error of immunity prevalence and risk with nationally representative electronic health record data. *J. Allergy Clin. Immunol.* 153:1704–1710. <https://doi.org/10.1016/j.jaci.2024.01.011>
22. Mayampurath, A., A. Ajith, C. Anderson-Smits, S.C. Chang, E. Brouwer, J. Johnson, M. Baltasi, S. Volchenboum, G. Devercelli, and C.E. Ciaccio. 2022. Early diagnosis of primary immunodeficiency disease using clinical data and machine learning. *J. Allergy Clin. Immunol. Pract.* 10:3002–3007.e5. <https://doi.org/10.1016/j.jaip.2022.08.041>
23. Yarmohammadi, H., L. Estrella, J. Doucette, and C. Cunningham-Rundles. 2006. Recognizing primary immune deficiency in clinical practice. *Clin. Vaccine Immunol.* 13:329–332. <https://doi.org/10.1128/CVI.13.3.329-332.2006>

Supplemental material

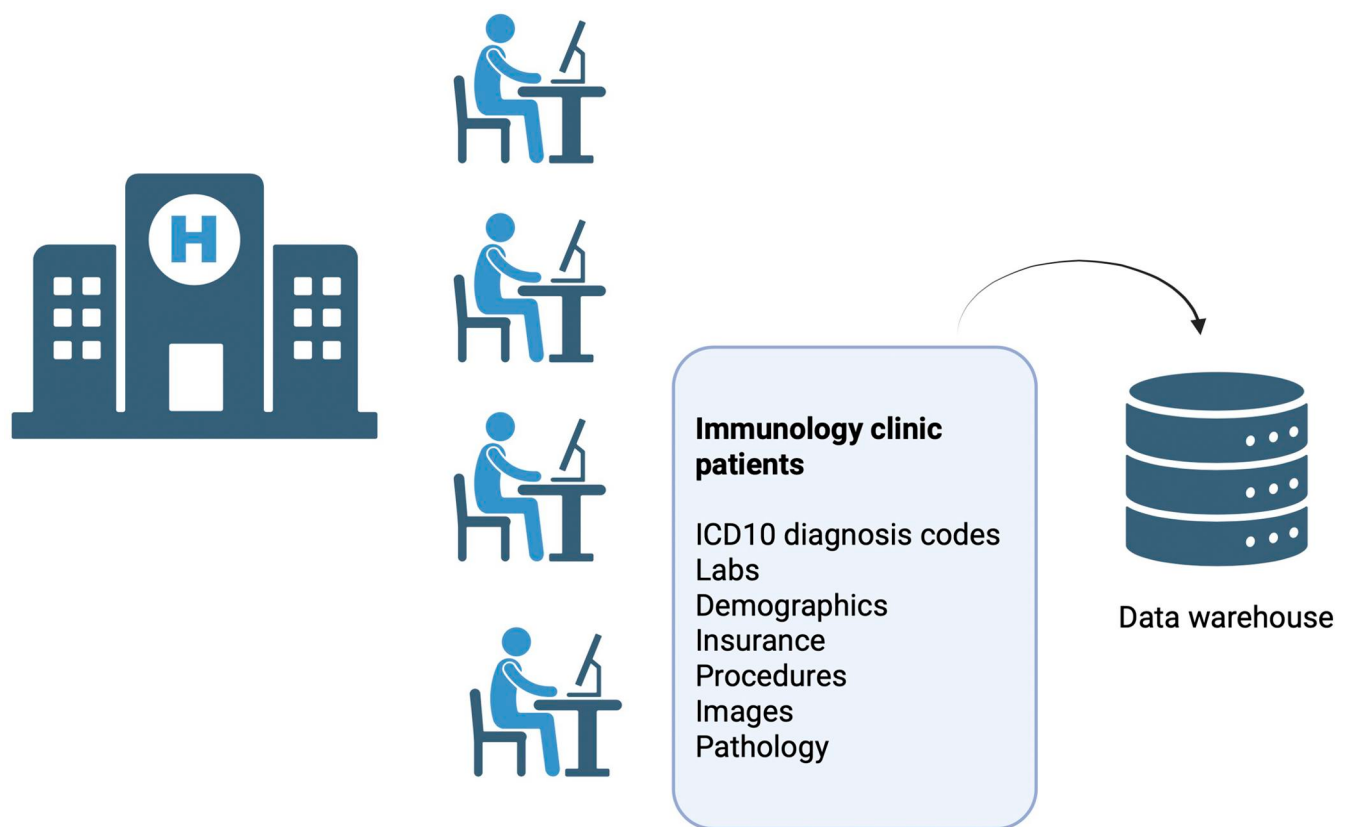


Figure S1. **USIDNET enrollment is based on immunology patients diagnosed and treated by clinical immunologists at different centers.** Discrete data fields are extracted from the defined patients in a common data model that requires no data entry other than that being provided through direct patient care. Although the cohort is defined as an immunology patient cohort, all visits and procedures are collected as are all medications and labs. Each event is time stamped with a day of life so timelines can be constructed.