Reporting of race and ethnicity in studies of artificial intelligence in pediatric urology: A secondary analysis of the AI-PEDURO online repository

Hyunwoong Harry Chae[1], David-Dan Nguyen[2,3,4], Adree Khondker[2,5], Mandy Rickard[5], Lauren Erdman[6,7], Andrew T. Gabrielson[8], Jin Kyu Kim[2,5,9], Jethro CC. Kwong[2], Brian Chun[10], Tariq Abbas[11], Nicolas Fernandez[12], Katherine Fischer[13], Lisette A. 't Hoen[14], Daniel T. Keefe[15], Hsin-Hsiao (Scott) Wang[16], John Weaver[17], Armando J. Lorenzo[2,5], Caleb P. Nelson[16]

[1]Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada; [2]Division of Urology, Department of Surgery, University of Toronto, Toronto, ON, Canada; [3]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; [4]Schwartz Reisman Institute for Technology and Society, University of Toronto, Toronto, ON, Canada; [5]Division of Urology, The Hospital for Sick Children, Toronto, ON, Canada; [6]Temerty Centre for AI Research and Education in Medicine, University of Toronto, Toronto, ON, Canada; [7]James M. Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital Medical Center and University of Cincinnati School of Medicine, Cincinnati, OH, USA; [8]James Buchanan Brady Urological Institute, Johns Hopkins University School of Medicine, Baltimore, MD, United States; [9]Division of Urology, Riley Children's Hospital, Indianapolis, IN, United States; [10]Division of Urology, Children's Hospital of Los Angeles, CA, United States; [11]Division of Urology, Sidra Medicine, Doha, Qatar; [12]Division of Pediatric Urology, Seattle Children's Hospital, Seattle, WA, United States; [13]Division of Urology, Children's Hospital of Philadelphia, Philadelphia, PA, United States; [14]Department of Urology, Sophia Children's Hospital, Erasmus University Medical Center, Rotterdam, The Netherlands; [15]Department of Urology, IWK Hospital, Halifax, NS, Canada; [16]Department of Urology, Boston Children's Hospital, Boston, MA, United States; [17]Department of Urology, Cleveland Clinic Children's Hospital, Cleveland, OH, United States

Cite as: Chae HH, Nguyen D-D, Khondker A, et al. Reporting of race and ethnicity in studies of artificial intelligence in pediatric urology: A secondary analysis of the AI-PEDURO online repository. *Can Urol Assoc J* 2026 March 30; Epub ahead of print. http://dx.doi.org/10.5489/cuaj.9541

Published online March 30, 2026

Corresponding author: Dr. Caleb P. Nelson, Department of Urology, Boston Children's Hospital, Boston, MA, United States; caleb.nelson@childrens.harvard.edu

***

ABSTRACT

Introduction: While an increasing number of artificial intelligence (AI) models are being developed in pediatric urology, the extent of race/ethnicity reporting among these studies is unclear. Our objective was to evaluate the inclusion and quality of race/ethnicity reporting in AI models in pediatric urology.

KEY MESSAGES

- Race and ethnicity reporting is absent among most published AI models in pediatric urology.
- Among studies that reported race/ethnicity data, the representation of minority racial/ethnic groups was poor.
- To ensure generalizability and fairness of AI models to be deployed in clinical pediatric urology settings, they must be evaluated for performance across diverse race/ethnicity subgroups and associated social determinants.

Methods: We conducted a secondary analysis of studies included in the AI in PEDiatric UROlogy (AI-PEDURO) collaborative living scoping review and repository. We examined racial/ethnic groups reported, their proportional representation, use of race/ethnicity as a predictor, conducting of stratified analyses by race/ethnicity, data collection methods, bias evaluation, and discussions of implications on equity.

Results: Of 81 studies in the AI-PEDURO repository, six (7.4%) reported race/ethnicity. Five studies included White and Black patients, representing 4824/7968 (60.5%) and 1377/7968 (17.3%) of the pooled cohort, respectively. Asian patients were included in three studies and represented 178/6861 (2.6%). Two studies reported Native Hawaiian or other Pacific Islander and Hispanic or Latino patients, representing 20/6704 (0.3%) and 1236/6704 (18.4%), respectively. One study included American Indian or Alaska Native patients, representing 69/6604 (1.0%). Mixed patients were included in three studies and represented 103/7711 (1.3%). Race/ethnicity was a predictor variable in 4/6 studies. None of these six studies conducted stratified analyses of model performance across race/ethnicity subgroups, reported race/ethnicity data collection methodologies, examined algorithmic biases, discussed implications on equity, or examined socioeconomic status or geographic residence.

Conclusions: Race/ethnicity reporting is poor in most AI studies in pediatric urology. Standardized reporting may help ensure fairness and generalizability of models across diverse pediatric urology populations.

INTRODUCTION

An increasing number of artificial intelligence (AI) models are being deployed in pediatric urology.[1,2] Previous studies of AI in pediatric urology have been conducted on a variety of topics, such as the prediction of outcomes in pediatric hydronephrosis, identifying risk factors of pediatric urinary tract infections (UTI), and improving accuracy in clinical assessments of hypospadias.[3–5] Multiple original research studies and systematic reviews have shown that AI holds promise for improving diagnostic and prognostic accuracy and supporting clinical decision-making.[6,7] However, the translation of AI into routine clinical practice (often referred to as the "AI chasm") remains limited due to challenges including insufficient external validation, difficulties integrating into existing workflows, regulatory barriers, and a lack of clinician trust.[8]

Generalizability and fairness are key concerns for AI models. Generalizability is the ability of an AI model to maintain adequate performance across diverse populations and clinical settings beyond those in which it was originally developed.[9] Fairness involves ensuring that model performance does not disproportionately benefit or harm particular subgroups.[10] The two concepts pose challenges that are especially relevant in pediatric urology, due to the specialty's small patient populations and unique pathologies and clinical presentations, which may limit model performance across diverse settings.[11] In existing studies, models trained and validated on narrow datasets may not reflect broader patient populations, limiting real-world applicability. Furthermore, biased algorithms can potentially reinforce existing disparities if they systematically underperform in specific marginalized populations.[12]

One key limitation in data quality that exists across AI studies in medicine is the omission of race and ethnicity variables in the development and/or evaluation of machine-learning (ML) models which is compounded by challenges in pediatric urology. Specifically, most existing studies are limited by small cohort sizes and incomplete inclusion of relevant clinical variables.[1] Furthermore, while many existing studies report the geographical distribution of their study centres, this alone does not meaningfully ensure generalizability of models across diverse racial/ethnic subgroups, especially since many countries are becoming increasingly multicultural.

The role of race and ethnicity in clinical medicine remains a contentious topic. As these are recognized as social constructs rather than biological determinants, some clinical organizations have recommended moving away from using race/ethnicity as variables that drive clinical decision-making.[13,14] However, while race/ethnicity may be flawed metrics for predicting biological and clinical outcomes, there is evidence suggesting that prevalence, risk factors, care delivery, and outcomes of pediatric urological conditions differ among racial and ethnic groups.[15,16] Regardless of the debate surrounding the appropriateness of inclusion of race/ethnicity data in AI models,[17] there is a strong case to be made for the need for increased standardized reporting of race and ethnicity in models being developed in pediatric urology and medicine. To reduce algorithmic bias and ensure predictions reflect real-world patient populations, performance of AI models must be evaluated across diverse population subgroups.

As such, given the importance of these variables for ensuring generalizability and fairness in AI models being used for clinical care in pediatric urology, we conducted a secondary analysis of the living scoping review and online repository developed by the AI in PEDiatric UROlogy (AI-PEDURO) collaborative, to characterize the reporting and use of race and ethnicity data in publications on AI in pediatric urology.[1]

METHODS

Study design

This is a secondary analysis of the living scoping review conducted by the AI-PEDURO collective. The original review searched 4 databases (MEDLINE (via Ovid), EMBASE (via Ovid), Scopus, and CINAHL (EBSCO)) without language restriction on July 1, 2025. Our search included studies where ML algorithms were used for clinical prediction and focused on a pediatric urology question. Biases of AI models were assessed using the APPRAISE-AI tool.[18] For reporting, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews (PRISMA-ScR).[19] Both the original protocol and scoping review have been published, and details of the search strategy, study selection criteria, data extraction, and model quality assessments are reported in the original publications.[1,2]

All 81 studies included in the original AI-PEDURO publication up to July 2025 were reviewed for the collection of race/ethnicity data.

Outcomes and evidence synthesis

The primary outcome was the number of AI studies in pediatric urology that have reported any race or ethnicity data. Secondary outcomes included whether method of race/ethnicity data collection was reported, the specific categories of race/ethnicity reported, whether race/ethnicity data was included in their ML predictor models, whether stratification analysis of model performance by racial/ethnic group was performed, if the study discussed potential racial/ethnic bias in ML algorithms and the implications of their model findings on health equity, methods utilized for and conclusions identified from their racial/ethnicity data analyses, and if the study reported social determinants associated with race/ethnicity, including variables related to socioeconomic status (e.g. insurance status, income, housing stability, education level, etc.), and geographic residence.

All outcomes were extracted by 2 independent reviewers. Discrepancies were resolved by consensus. These characteristics were reported using descriptive statistics. All analyses were performed in R 4.0.4 (R Foundation for Statistical Computing, Vienna, Austria). Descriptive variables were tabulated and compared, qualitatively, and numerical variables such as proportion or counts were pooled.

RESULTS

Race/ethnicity reporting

Overall, a total of 81 publications were studied in this review. Publication years ranged from 1998 to 2025, with 66/81 (81.4%) of studies being published after 2019.

Of the total group, 6 studies (7.4%) included race or ethnicity as a variable (Table 1; study numbers in brackets correspond only to those included in Table 1), all of which were conducted in the U.S. No study reported the methodology utilized to collect race/ethnicity data. White patients were the most frequently reported group, appearing in 5/6 studies reporting race or ethnicity. White patients also comprised the majority of patients in these studies: 482/607 (79.4%) [1], 368/500 (73.6%) [2], 68/157 (43.3%) [4], 53/100 (53.0%) [5], and 3853/6604 (58.3%) [6] patients in these 5 study cohorts. In total across these 5 studies, White patients represented 4824/7968 (60.5%) participants.

Black patients were included as a racial group in 5/6 studies that reported race or ethnicity. They represented 27/607 (4.4%) [1], 52/500 (10.4%) [2], 42/157 (26.8%) [4], 21/100 (21.0%) [5], 1235/6604 (18.7%) [6] patients in these 5 study cohorts. In total across these 5 studies, Black patients represented 1377/7968 (17.3%) participants.

Asian patients were included as a racial group in 3/6 of studies that reported race or ethnicity. They represented 7/157 (4.5%) [4], 4/100 (4.0%) [5], 167/6604 (2.5%) [6] patients in these 3 study cohorts. In total across these 3 studies, Asian patients represented 178/6861 (2.6%) participants.

Hispanic or Latino ethnicity was included as an ethnic group in 2/6 studies that reported race or ethnicity. They represented 6/100 (6.0%) [5] and 1230/6604 (18.6%) [6] patients in these 2 study cohorts. In total across these 2 studies, Hispanic or Latino patients represented 1236/6704 (18.4%) participants.

Native Hawaiian or other Pacific Islander was included as a racial group in 2/6 studies that reported race or ethnicity. They represented 1/100 (1.0%) [5] and 19/6604 (0.3%) [6] patients in these 2 study cohorts. In total across these 2 studies, Native Hawaiian or Other Pacific Islander patients represented 20/6704 (0.3%) participants.

American Indian or Alaska Native was included as a racial group in 1/6 studies that reported race or ethnicity. They represented 69/6604 (1.0%) [6] patients in this study cohort.

Mixed race was included as a group in 3/6 studies that reported race or ethnicity. They represented 40/607 (6.6%) [1], 37/500 (7.4%) [2], 26/6604 (0.4%) [6] patients in these 3 study cohorts. In total across these 3 studies, Mixed race patients represented 103/7711 (1.3%) participants.

Use of race/ethnicity data

Four of the 6 studies included race/ethnicity as a predictor variable in their models. Three of these 4 studies identified significant findings based on analysis of race/ethnicity data, as Bertsimas et al. [1] concluded that race plays an essential factor in risk stratification for recurrent

UTI, Estrada et al. [2] identified that White patients were associated with a higher risk of recurrent UTI, and Suh [6] identified that Black race was a risk factor for chronic rejection-caused graft failure in pediatric kidney transplant recipients 15 years post-transplant. None of the 6 studies conducted any stratified analyses of model performance across race/ethnicity subgroups, addressed potential algorithmic bias related to race/ethnicity, discussed whether the findings from their AI models had implications for health equity across racial/ethnic groups, nor reported variables related to socioeconomic status or geographic residence.

DISCUSSION

In this secondary analysis of the AI-PEDURO living review, only 6 of 81 (7.4%) studies of AI models in pediatric urology reported race or ethnicity. In studies reporting race/ethnicity data, White patients were the most prevalent, while several racial/ethnic groups were significantly underrepresented with <3% of the overall cohort. This raises concerns for the generalizability and fairness of currently developed AI models in pediatric urology, presenting a potential barrier for broad and equitable deployment in clinical practice.

Our finding of limited reporting of race/ethnicity in studies of AI in pediatric urology aligns with previous studies of AI models in other fields. Multiple prior systematic reviews of AI studies across medicine have reported significant under-reporting of race/ethnicity.[20,21] Underreporting of racial/ethnic minorities is also prevalent across non-AI-related studies in urology, including in randomized controlled trials of conditions such as benign prostatic hyperplasia and urologic cancers.[22] Collectively, these findings highlight that underreporting of race/ethnicity is a pervasive issue that spans across AI research and urology literature.

Among studies that reported race or ethnicity, most participants were White, with minority racial and ethnic groups markedly underrepresented. This finding is consistent with prior reviews of AI studies in mammogram interpretation and general medical diagnoses, both of which reported predominantly White study cohorts with underrepresentation of minority racial and ethnic groups.[23,24] Ideally, future AI models in pediatric urology should make a concerted effort to include comparable numbers of patients across different racial/ethnic groups. However, as real-world populations do not have equal distribution across racial/ethnic groups, this goal would require large sample sizes, making it logistically challenging in practice. Nonetheless, increasing reporting of race/ethnicity in AI studies in pediatric urology and ensuring training data have equitable representation of different racial/ethnic groups is crucial to ensure generalizability. Machine learning algorithms tend to overfit to their training data, which entails learning various relationships present in the training data.[25] If race/ethnicity is not reported or certain racial/ethnic groups are underrepresented, the model may fail to capture clinically relevant patterns for specific subpopulations. Consequently, the model's performance when applied in patient populations different from its training data may be limited. Such limitations are already evident in pediatric urology literature. For example, there have already been 5 published pediatric urology AI models that predict the need for urological surgery in children based on institutional training data without available race/ethnicity data.[26–30] Without proper evaluation of

performance across racial/ethnic subgroups, these models could inadvertently exacerbate disparities for this subjective clinical decision.

None of the included studies in this review conducted stratified analyses of model performance across race/ethnicity subgroups or discussed potential bias in their models, highlighting that current AI studies in pediatric urology are not conducting sufficient bias assessments. Fairness in AI is achieved when algorithmic decision-making does not favour or discriminate against a specific group based on protected attributes, including race and ethnicity.[25] Biases must be identified and mitigated to preserve algorithmic fairness. Lack of bias assessment in AI models is prevalent in medicine beyond just pediatric urology. According to a prior scoping review of bias instances in clinical AI models, only 12% of existing clinical ML studies performed any algorithmic bias assessments, and 75% of AI models showed bias against a specific sociodemographic group.[31] Increased utilization of AI in healthcare settings has already been associated with the exacerbation of racial disparities in minority populations.[32] These findings are concerning, as without robust bias assessment of AI models (e.g., evaluating performance across racial/ethnic subgroups) before clinical deployment, existing inequities may be perpetuated, potentially leading to ongoing disparities in the care of pediatric urology patients belonging to minority racial/ethnic groups.[18]

While the exact association between race/ethnicity and health outcomes may be an ongoing debate, in pediatric urology, some existing studies have demonstrated that children of different race/ethnicity are associated with differences in disease prevalence, presentation, and outcomes of urological conditions. For example, African American children have been reported to have younger age of pediatric pyeloplasty compared to White children.[33,34] Other studies have reported associations between Black patients and higher rates of morbidity after urologic surgery, differences in the prevalence of congenital anomalies, and variations in symptom severity of lower urinary tract dysfunction.[15,35,36]

Conversely, arguments against the inclusion of race/ethnicity in clinical decision-making tools and AI models are also prevalent in the literature. Fundamentally, race and ethnicity are recognized as social constructs rather than biologically defined categories. As stand-alone variables, they are flawed metrics that are imperfect in capturing complex underlying social, environmental, and genetic factors that may influence clinical outcomes or model performance. As such, studies in the field of genomics have recently been recommended to move away from the use of race, ethnicity, and geographic origin to be used as proxies for genetic ancestry groups.[37] Additionally, multiple clinical organizations have recently advocated for the transition away from race-based medicine. In 2021, the American Academy of Pediatrics (AAP) retired their 2011 UTI guidelines, which included that patients who are not Black should have increased pre-test probability of UTI for febrile infants 2–24 months old. The AAP determined that this algorithm may result in underdiagnosing UTI in Black children, and that race should not be viewed as a risk factor that predicts disease, disease severity and disability but rather as a risk marker of bias, discrimination and vulnerability.[13] Similarly, in 2021, the National Kidney

Foundation (NKF) and American Society of Nephrology (ASN) recommended transitioning away from race-adjusted eGFR equations, citing that race is a social construct rather than a biological variable, and its use in eGFR can cause delayed chronic kidney disease (CKD) diagnoses and subsequent referrals, reinforcing inequities and introducing systematic bias into the care of minority patient populations.[14] Despite these arguments, some studies have expressed that this transition away from race-based medicine is not without its own associated risks and harms, as they have reported that models developed without inclusion of race/ethnicity data had lower performance and accuracy when applied in certain clinical contexts.[38,39]

These diverse viewpoints made by different stakeholders highlight an important underlying point, that the proper role of race/ethnicity data in AI, research, and clinical medicine remains uncertain. While the appropriateness of including race/ethnicity data in AI models that drive clinical decision-making may continue to be debated, we believe that standardized reporting of race/ethnicity data is essential and necessary when developing AI models for application in pediatric urology. To ensure equitable performance across different racial/ethnic populations, AI models must be evaluated in diverse racial/ethnic subgroups. Even if race/ethnicity is not ultimately included in AI models being directly applied to clinical scenarios and used to drive decision-making, without reporting this data in the first place, the generalizability and fairness of these models cannot be tracked nor evaluated to begin with. For these reasons, multiple studies have already advocated for increasing reporting of race/ethnicity data in research studies across AI, urology, and medicine broadly.[20–22]

Furthermore, our review identified that none of the six included studies reporting race/ethnicity reported associated variables related to socioeconomic status or geographic residence. Race and ethnicity are associated with various social determinants of health beyond just socioeconomic status and geographic residence, such as immigration status, language barriers, patient trust in the healthcare system, and systemic inequities in access to care, which all may potentially confound the influence of race/ethnicity on health outcomes.[40] This further highlights the need for future AI models to be carefully evaluated, not only across racial/ethnic subgroups, but also with consideration for other intersecting social determinants of health.

Researchers may face practical and logistical barriers in collecting and reporting these data. One report by the Council of State and Territorial Epidemiologists in 2022 concluded that based on survey results, commonly reported barriers to race/ethnicity data collection by public health agencies in the U.S. include legal/regulatory limitations, information systems limitations, insufficient guidance, requirements and standards for data collection, limited resources and staffing, lack of reporting from data submitters (clinicians, hospitals, pharmacies, etc.), and patients' hesitance to share information about their race/ethnicity.[41]

Our study has limitations. First, none of the included studies described the methodology they used to collect race/ethnicity data, introducing uncertainty regarding the validity and consistency of these variables. Second, all 6 studies that reported race/ethnicity data were conducted in the U.S., potentially limiting the applicability of our findings to international

settings with different demographic profiles and healthcare systems. Finally, as the AI-PEDURO is a living scoping review, and AI models in pediatric urology are rapidly evolving, the trends regarding race/ethnicity reporting in publications on this topic may change over time.

CONCLUSIONS

Only 7.4% of AI studies in pediatric urology reported race or ethnicity data, with minority racial and ethnic groups markedly underrepresented. None of these studies described data collection methods, evaluated algorithmic bias, discussed implications for health equity, or reported on socioeconomic status and geographic residence. Our findings support the broader principle that race/ethnicity and associated social determinants, should always be considered amongst all clinical studies, in order to appropriately describe populations and measure disparities. In AI studies, these variables are especially essential to ensure generalizability and fairness of models. While the use of these variables as predictors in AI models continues to be debated, our findings emphasize that consistent reporting is necessary for evaluation of model performance across diverse population subgroups, and ultimately help prevent the perpetuation of existing disparities in pediatric urology and medicine more broadly.

REFERENCES

1. Khondker A, Kwong JC, Ahmad I, et al. A living scoping review and online repository of artificial intelligence models in pediatric urology: Results from the AI-PEDURO collaborative. *J Pediatr Urol* 2025;21:765-72. https://doi.org/10.1016/j.jpurol.2025.01.035
2. Khondker A, Kwong JCC, Rickard M, et al. AI-PEDURO - Artificial intelligence in pediatric urology: Protocol for a living scoping review and online repository. *J Pediatr Urol* 2025;21:532-8. https://doi.org/10.1016/j.jpurol.2024.10.003
3. Bertsimas D, Li M, Estrada C, et al. Selecting children with vesicoureteral reflux who are most likely to benefit from antibiotic prophylaxis: Application of machine learning to RIVUR. *J Urol Wolters Kluwer Health* 2021:1170-1178. https://doi.org/10.1097/JU.0000000000001445
4. Fernandez N, Lorenzo AJ, Rickard M, et al. Digital pattern recognition for the identification and classification of hypospadias using artificial intelligence vs experienced pediatric urologist. *Urology* 2021;147:264-9. https://doi.org/10.1016/j.urology.2020.09.019
5. Khondker A, Kwong JCC, Chancy M, et al. Predicting obstruction risk using common ultrasonography parameters in paediatric hydronephrosis with machine learning. *BJU Int* 2024;133:79-86. https://doi.org/10.1111/bju.16159
6. Bouhadana D, Lu XH, Luo JW, et al. Clinical applications of machine learning for urolithiasis and benign prostatic hyperplasia: A systematic review. *J Endourol* 2023;37:474-94. https://doi.org/10.1089/end.2022.0311
7. Kwong JCC, Wu J, Malik S, et al. Predicting non-muscle invasive bladder cancer outcomes using artificial intelligence: a systematic review using APPRAISE-AI. *npj Digit Med* 2024;7:1-11. https://doi.org/10.1038/s41746-024-01088-7
8. Kwong JCC, Erdman L, Khondker A, et al. The silent trial - the bridge between bench-to-bedside clinical AI applications. *Front Digit Heal* 2022;4:929508. https://doi.org/10.3389/fdgth.2022.929508
9. Goetz L, Seedat N, Vandersluis R, et al. Generalization-a key challenge for responsible AI in patient-facing clinical applications. *npj Digit Med* 2024;7:1-4. https://doi.org/10.1038/s41746-024-01127-3
10. Ueda D, Kakinuma T, Fujita S, et al. Fairness of artificial intelligence in healthcare: review and recommendations. *Jpn J Radiol* 2023;42:3. https://doi.org/10.1007/s11604-023-01474-3
11. Khondker A, Kwong JCC, Malik S, et al. The state of artificial intelligence in pediatric urology. *Front Urol* 2022;2:1024662. https://doi.org/10.3389/fruro.2022.1024662
12. Chin MH, Afsar-Manesh N, Bierman AS, et al. Guiding principles to address the impact of algorithm bias on racial and ethnic disparities in health and health care. *JAMA Netw Open* 2023;6:E2345050. https://doi.org/10.1001/jamanetworkopen.2023.45050
13. Beers LS, Szilagyi M, Goza SH, et al. AAP perspective: Race-based medicine. *Pediatrics* 2021;148. https://doi.org/10.1542/peds.2021-053829
14. Delgado C, Baweja M, Crews DC, et al. A Unifying approach for GFR estimation: Recommendations of the NKF-ASN task force on reassessing the inclusion of race in

diagnosing kidney disease. *Am J Kidney Dis* 2022;79:268-288.e1. https://doi.org/10.1053/j.ajkd.2021.08.003
15. Ko JS, Corbett C, Fischer KM, et al. Impact of race, gender, and socioeconomic status on symptom severity at time of urologic referral. *J Racial Ethn Heal Disparities* 2023;10:1735-44. https://doi.org/10.1007/s40615-022-01357-9
16. Klein JB, Nguyen CT, Saffore L, et al. Racial disparities in urologic health care. *J Natl Med Assoc* 2010;102:108-18. https://doi.org/10.1016/S0027-9684(15)30498-3
17. Vyas DA, Eisenstein LG, Jones DS. The race-correction debates - progress, tensions, and future directions. *N Engl J Med* 2025;393:1029-36. https://doi.org/10.1056/NEJMms2506241
18. Kwong JCC, Khondker A, Lajkosz K, et al. APPRAISE-AI tool for quantitative evaluation of AI studies for clinical decision support. *JAMA Netw Open* 2023;6:2335377. https://doi.org/10.1001/jamanetworkopen.2023.35377
19. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Ann Intern Med* 2018;169:467-73. https://doi.org/10.7326/M18-0850
20. Elmahdy M, Sebro R. Sex, ethnicity, and race data are often unreported in artificial intelligence and machine learning studies in medicine. *Intell Med* 2023;8:100113. https://doi.org/10.1016/j.ibmed.2023.100113
21. Bozkurt S, Cahan EM, Seneviratne MG, et al. Reporting of demographic data and representativeness in machine learning models using electronic health records. *J Am Med Informatics Assoc* 2020;27:1878-84. https://doi.org/10.1093/jamia/ocaa164
22. Nguyen AL V, Moustafa M, Nguyen DD, et al. Absence of race/ethnicity reporting in clinical trials of true minimally invasive surgical therapies for the treatment of benign prostatic hyperplasia. *Urology* 2025;196:300-8. https://doi.org/10.1016/j.urology.2024.09.043
23. Miyawaki IA, Banerjee I, Batalini F, et al. Global disparities in artificial intelligence-based mammogram interpretation for breast cancer: A scientometric analysis of representation, trends, and equity. *Eur J Cancer* 2025;220:115394. https://doi.org/10.1016/j.ejca.2025.115394
24. Fitzsimmons L, Dewan M, Dexheimer JW. Diversity in machine learning: A systematic review of text-based diagnostic applications. *Appl Clin Inform* 2022;13:569-82. https://doi.org/10.1055/s-0042-1749119
25. Huang J, Galal G, Etemadi M, et al. Evaluation and mitigation of racial bias in clinical machine learning models: Scoping review. *JMIR Med Informatics* 2022;10:e36388. https://doi.org/10.2196/36388
26. Blum ES, Porras AR, Biggs E, et al. Early detection of ureteropelvic junction obstruction using signal analysis and machine learning: A dynamic solution to a dynamic problem. *J Urol* 2018;199:847-52. https://doi.org/10.1016/j.juro.2017.09.147
27. Erdman L, Skreta M, Rickard M, et al. Predicting obstructive hydronephrosis based on ultrasound alone. In: Lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Vol 12263 LNCS. *Springer Science and Business Media Deutschland GmbH* 2020:493-503. https://doi.org/10.1007/978-3-030-59716-0_47

28. Guan Y, Wen P, Li J, et al. A DWT-Utilized classifier for UPJO diagnosis using ultrasound images. In: ICNSC 2022 - Proceedings of 2022 IEEE international conference on networking, sensing and control: Autonomous intelligent systems. I*nstitute of Electrical and Electronics Engineers Inc* 2022. https://doi.org/10.1109/ICNSC55942.2022.10004150
29. Seçkiner I, Seçkiner SU, Bayrak Ö, et al. Use of artificial neural networks in the management of antenatally diagnosed ureteropelvic junction obstruction. *J Can Urol Assoc* 2011;5. https://doi.org/10.5489/cuaj.10043
30. Lorenzo AJ, Rickard M, Braga LH, et al. Predictive analytics and modeling employing machine learning technology: The next step in data sharing, analysis, and individualized counseling explored with a large, prospective prenatal hydronephrosis database. *Urology* 2019;123:204-9. https://doi.org/10.1016/j.urology.2018.05.041
31. Colacci M, Huang YQ, Postill G, et al. Sociodemographic bias in clinical machine learning models: A scoping review of algorithmic bias instances and mechanisms. J *Clin Epidemiol* 2025;178:111606. https://doi.org/10.1016/j.jclinepi.2024.111606
32. Haider SA, Borna S, Gomez-Cabello CA, et al. The algorithmic divide: A systematic review on AI-driven racial disparities in healthcare. *J Racial Ethn Heal Disparities* Published online 2024. https://doi.org/10.1007/s40615-024-02237-0
33. Khondker A, Dhalla R, Nguyen DD, et al. Does marginalization status and distance from the hospital impact the timing of pediatric pyeloplasty in a universal access-to-care system? *Can Urol Assoc J* 2025;19:249-54. https://doi.org/10.5489/cuaj.9133
34. Nelson CP. Evidence of variation by race in the timing of surgery for correction of pediatric ureteropelvic junction obstruction. *J Urol* 2007;178:1463-8. https://doi.org/10.1016/j.juro.2007.05.167
35. Chu DI, Canning DA, Tasian GE. Race and 30-day morbidity in pediatric urologic surgery. *Pediatrics* 2016;138. https://doi.org/10.1542/peds.2015-4574
36. Lloyd JC, Wiener JS, Gargollo PC, et al. Contemporary epidemiological trends in complex congenital genitourinary anomalies. *J Urol* 2013;190:1590-5. https://doi.org/10.1016/j.juro.2013.04.034
37. Feero WG, Steiner RD, Slavotinek A, et al. Guidance on use of race, ethnicity, and geographic origin as proxies for genetic ancestry groups in biomedical publications. *JAMA* 2024;331:1276-8. https://doi.org/10.1001/jama.2024.3737
38. Khor S, Haupt EC, Hahn EE, et al. Racial and ethnic bias in risk prediction models for colorectal cancer recurrence when race and ethnicity are omitted as predictors. *JAMA Netw Open* 2023;6:E2318495. https://doi.org/10.1001/jamanetworkopen.2023.18495
39. Ladin K, Cuddeback J, Duru OK, et al. Guidance for unbiased predictive information for healthcare decision-making and equity (GUIDE): Considerations when race may be a prognostic factor. *npj Digit Med* 2024;7:1-12. https://doi.org/10.1038/s41746-024-01245-y
40. Walker RJ, Strom Williams J, Egede LE. Influence of race, ethnicity and social determinants of health on diabetes outcomes. *Am J Med Sci* 2016;351:366-373. https://doi.org/10.1016/j.amjms.2016.01.008
41. Beaulieu B. Addressing gaps in public health reporting of race and ethnicity data for COVID-19: Findings and recommendations among 45 state & local health departments.; 2022.

FIGURES AND TABLES

Table 1. Summary of publications that reported race/ethnicity as a variable

Study*	Race/ethnicity reported	Method of race/ethnicity data collection reported	White (n, %)	Black (n, %)	Asian (n, %)	Native Hawaiian or other Pacific Islander (n, %)	American Indian or Alaska Native (n, %)	Hispanic or Latino ethnicity (n, %)	Mixed (n, %)	Other (n, %)
Bertsimas 2021[1]	Yes	No	482/607 (79.4%)	27/607 (4.4%)	NA	NA	NA	NA	40/607 (6.6%)	48/607 (7.9%)
Estrada 2019[2]	Yes	No	368/500 (73.6%)	52/500 (10.4%)	NA	NA	NA	NA	37/500 (7.4%)	35/500 (7.0%)
Weaver 2023[**3]	Yes	No	NA	NA	NA	NA	NA	NA	NA	NA
Yin 2020[4]	Yes	No	68/157 (43.3%)	42/157 (26.8%)	7/157 (4.5%)	NA	NA	NA	NA	40/157 (25.5%)
Zheng 2019[5]	Yes	No	53/100 (53.0%)	21/100 (21.0%)	4/100 (4.0%)	1/100 (1.0%)	NA	6/100 (6.0%)	NA	NA
Suh 2025[***6]	Yes	No	3853/6604 (58.3%)	1235/6604 (18.7%)	167/6604 (2.5%)	19/6604 (0.3%)	69/6604 (1.0%)	1230/6604 (18.6%)	26/6604 (0.4%)	NA
Total pooled percentage	6/6 (100%) of studies	0/6 (0%) of studies	4824/7968 (60.5%)	1377/7968 (17.3%)	178/6861 (2.6%)	20/6704 (0.3%)	69/6604 (1.0%)	1236/6704 (18.4%)	103/7711 (1.3%)	123/1264 (9.7%)

*Studies included in Table 1 are cited in the text using bracketed study numbers. **Weaver 2023 stated that race was included as clinical variable in a random forest model but did not provide any breakdown of race/ethnicity data. ***Suh 2025 only provided % of patients for each race/ethnicity category and did not provide raw numbers of patients. Raw counts were reverse calculated from reported percentages by multiplying each percentage by the study sample size and rounding to the nearest whole number. See online Appendix (at *cuaj.ca*) for table references.

Table 1 (cont'd). Summary of publications that reported race/ethnicity as a variable								
Unknown (n, %)	**Race/ethnicity included as predictor in model**	**Stratified analysis of model performance across race/ethnicity subgroups performed**	**Discussed racial/ethnic bias in models**	**Discussed health equity implications across racial/ethnic groups based on model findings**	**Methods for race/ethnicity analyses**	**Significant findings from race/ethnicity data analyses**	**Socioeconomic status-related variable reported**	**Geographic residence-related variable reported**
NA	Yes	No	No	No	Race included as a predictor of outcome in final model.	"Race, gender and VUR grade played essential roles in rUTI risk stratification for both CAP and placebo models which was consistent with current literature."	No	No
8/500 (1.6%)	Yes	No	No	No	Race included as a predictor of outcome in final model.	"Our model also showed a higher risk of rUTI associated VUR in White and female children, which was consistent	No	No

						with reported data."		
NA	Yes	No	No	No	Race included in a random forest model that used clinical variables to predict outcome. Study did not provide a breakdown of racial/ethnic data.	None.	No	No
NA	No	No	No	No	Race included in demographic characteristics but was not included as a predictor in the model.	None.	No	No
21/100 (21.0%)	No	No	No	No	Race included in demographic characteristics only but not included as a predictor in the model.	None.	No	No
NA	Yes	No	No	No	Race included as a predictor	Black race was a risk	No	No

					of outcome in final model (Random Forest).	factor for chronic rejection-caused graft failure in pediatric kidney transplant recipients 15 years post-transplant.		
29/600 (4.8%)	4/6 (66.7%) of studies	0/6 (0%) of studies	0/6 (0%) of studies	0/6 (0%) of studies	NA	NA	0/6 (0%) of studies	0/6 (0%) of studies

*Studies included in Table 1 are cited in the text using bracketed study numbers. **Weaver 2023 stated that race was included as clinical variable in a random forest model but did not provide any breakdown of race/ethnicity data. ***Suh 2025 only provided % of patients for each race/ethnicity category and did not provide raw numbers of patients. Raw counts were reverse calculated from reported percentages by multiplying each percentage by the study sample size and rounding to the nearest whole number. See online Appendix (at *cuaj.ca*) for table references.