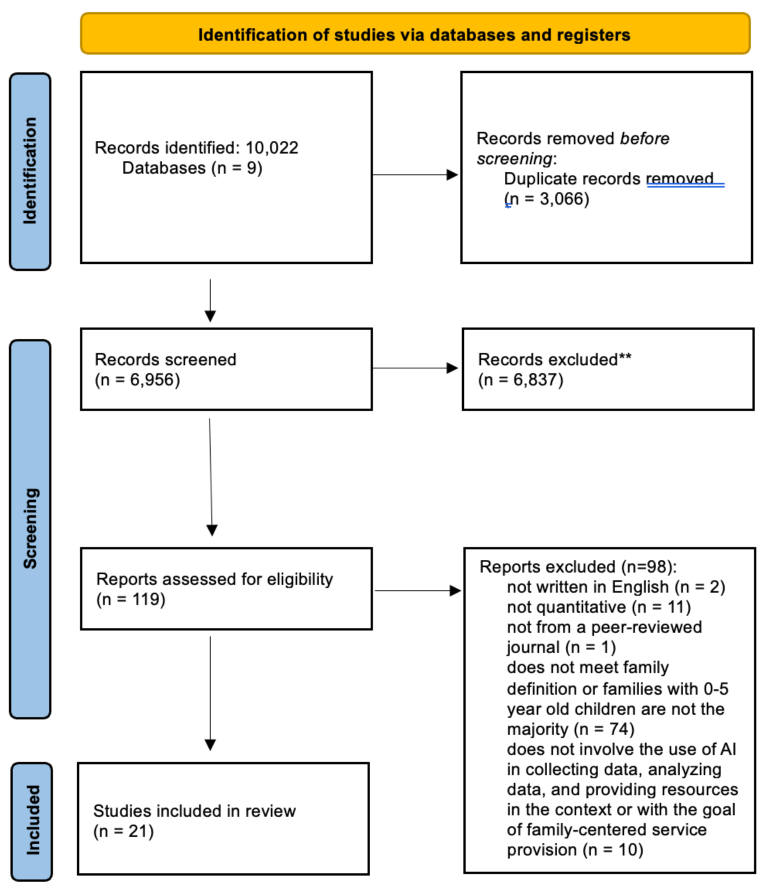
Introduction

A growing body of population research with a focus on families is exploring the use of artificial intelligence (AI) technologies and methodologies to understand family functioning such as maternal and child health (MCH) and other health matters relevant to families raising young children (Lopez-Larrosa et al., 2023; Mlandu et al., 2023; Pan et al., 2017; Sun, 2024). The application of AI in the field of demography and population research is promising because these technologies can potentially inform policy decisions to support families with young children. However, there is limited understanding of the evidence base across studies, and key questions remain unanswered, such as the nature of the AI tools being utilized and the methods employed, the consolidated outcomes of the studies, and the ethical application of AI.

A number of review studies have been conducted to more systematically examine the use of AI in research involving individuals expecting new babies (e.g., expectant mothers) or families raising young children (Bertini et al., 2022; Du et al., 2023; Khan et al., 2022; Krakouer et al., 2021; Robila & Robila, 2020; Triantafyllidis et al., 2020). However, there are several limitations to this evidence base including: (1) using broad definitions of family (e.g., that not only includes children and parents, but also veterans and the elderly) which makes it challenging to identify common patterns across studies; (2) focusing on singular topics, models, and sources of data; and (3) lacking a systematic and critical examination of the ethnical and fair use of AI.

To address limitations in the literature, this scoping review systematically examined the literature to identify AI use in studying families, especially those with young children. It also examined whether studies considered and clearly articulated a statement of ethically and fairly employing AI. This study makes an important contribution to and advances the field by elucidating the role of AI in promoting the study populations involving families with young children. It provides timely evidence on the types of AI research conducted in the field while also identifying gaps in the literature, along with future recommendations.

FIGURE 1. PRISMA flowchart



Methods

The current scoping review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews (PRISMA-ScR; (Tricco et al., 2018) and preregistered at Open Science Framework for transparency and reproducibility purposes: https://www.doi.org/10.17605/OSF.IO/5A92D. Nine databases were comprehensively searched (e.g., PubMed, Excerpta Medica dataBASE, PsycINFO, Cochrane Library, Psychology and Behavioral Sciences Collection, Social Sciences Abstracts, and Web of Science Core Collection). Eligible studies were: (1) peer-reviewed journal articles; (2) published between 2014–2024; written in English; (3) involved the use of AI; (4) included families raising young children 0–5 years; and (6) were quantitative in analysis.

As shown in Figure 1 (PRISMA Flowchart), a total of 10,002 studies were initially identified and screened, resulting in a total of 21 final studies. Data were extracted based on publication characteristics (e.g., author, year, study aim), participant characteristics (e.g., child age, geographic location), study characteristics (e.g., sample size, study design, domain of AI use, ethnical statements concerning the use of AI). After data extraction, a narrative synthesis was conducted in accordance with Popay et al. (2005), and studies were grouped into three domains of AI use: (a) collection of family demography data; (b) analysis of family demography data; and (c) delivery of services to families.

Findings

Preliminary Results

The final 21 studies were published between 2017 and 2024, with nearly 60% (*n*=12) of the studies focusing on family samples from low- and middle-income countries (LMICs; e.g. African countries such as Ethiopia, Rwanda, Zanzibar). Although data sources varied across studies, the majority of used national survey data (*n*=11; 52%). Studies examined a wide variation within MCH outcomes (e.g. maternal healthcare utilization, infant mortality, infant health, mothers' adherence to exclusive breastfeeding recommendations, child obesity, maternal mortality, maternal adverse birth outcomes, family violence, home visitation services for child maltreatment prevention). Most studies (*n*=15; 71%) utilized more than three machine learning models for robust data analysis, and the most used model was random forest (*n*= 12; 57%).

Main Results

The main results demonstrating AI use in population research involving families with young children are shown in Table 1. All studies used AI for data analysis purposes, either identifying events of interest (*n*=11; 52%) or finding the most important predictors (*n*=16; 76%). Although most studies focused on specific MCH outcomes, Lee et al. (2022) was the only study that more broadly focused on an outcome that could impact the whole family (i.e., norovirus). A few studies (*n*=2 studies or 10%) utilized AI for assessing treatment effects specifically through the use of double machine learning, which is being increasingly applied to enhance causal inference (Peet et al., 2023, 2024). Most of the studies did not consider, and thus did not mention, the ethical use of AI (e.g., perpetration of bias such as systemic racism and discrimination, data privacy, algorithmic fairness, accountability, transparency or the "black box" problem, and appropriateness of AI, and reproducibility of research). In fact, only one study specifically addressed ethical concerns related to using AI in their studies (Ahn et al., 2024), discussing potential biases in AI models that could lead to unfair treatment or discrimination.

Discussion

The current scoping review systematically reviewed the population literature to examine how AI is being leveraged to study the health and well-being of families raising young children. In summary, the evidence base that employs AI to population research involving families with young children is limited in size and scope, with most studies using AI for data analysis purposes (i.e., identifying most important predictors of MCH outcomes using random forest). AI models focused on predicting events of interest may be useful for family policymakers in prevention work, particularly those promoting public health among MCH issues. Various events of interest, such as infant mortality, child maltreatment, and intimate partner violence in families, would benefit from family services that can address events prior to intervention. For example, norovirus outbreak can be deterred with early and proper social distancing measures by health authorities and schools, and infant mortality can be reduced by encouraging maternal enrollment in antenatal care.

Although some studies found several identical predictors of MCH outcomes across geographical locations and cultural context, others found quite disparate predictors even if outcomes were similar or nearly identical. Regarding variations in outcomes across studies, this may be attributed to geographic and cultural differences. Future research in this area should consider and account for countries' differences in historical, social, and environmental factors that may help explain differences in outcomes. Importantly, most studies reviewed in the current review did not include comprehensive ethical statements regarding the use of AI, and even when they did, these statements were often limited to addressing general data limitations and potential biases without delving deeply into the ethical implications, considering the context of AI application. Future research in the field of population and demography that employs machine learning and other AI-facilitated tools should provide statements on ethical AI use. At the minimum, studies should engage in thoughtful discussion of how their use of AI may be perpetuating systemic racism, discrimination, and biases against marginalized populations and detail measures in place for prevention.

TABLE 1. Main results from analyzing the final set of articles (*N* = 21)

Reference	Results
	AI use to identify events of interest (prediction)
Ahn et al. (2024)	• Relative to the baseline model's (i.e., existing assessment tool) performance of 46.2%, the predictive risk model (PRM) accurately identified 75.3%–84.1% of children experiencing substantiated child maltreatment and thus in need of home-visiting services.
Demsash (2023)	• Relative to the naïve Bayes, logistic regression, J48 classifier, and AdaBoost models, the random forest model performed the best and predicted child mortality before the age of 5 years old at 93.9% accuracy.
Fredriksson et al. (2022)	• Relative to logistic regression, LASSO regularized logistic regression, random forest, and ANN models, the random forest model performed the best, identifying accurately 74.4% of mothers' childbirth delivery in the home over childbirth delivery in a health facility.
Hazlett et al. (2023)	• Relative to the baseline model (i.e., indicators of wealth as the only predictors of infant mortality), which only captured 24% of infant mortality in the top quintile of risk, the elastic-net logit, random forest, extreme gradient boosting, kernel regularized least squares models with additional variables (e.g., previous death of a sibling in the first year, child's gender, rural/urban indicator, mothers' age and education) predicted infant mortality risk more accurately and captured 34% of infant mortality in the top quintile of risk. The elastic-net logit, random forest, extreme gradient boosting, and kernel regularized least squares models had similar average performances.
Lee et al. (2022)	• Relative to the support vector machine (SVM), multilayer perceptron (MLP), random forest, gradient boosting (GB), and gated recurrent unit (GRU) models, the long short-term memory (LSTM) model performed the best, predicting with 92.5% accuracy weekly norovirus outbreaks.
Mondal et al. (2023)	• Relative to logistic regression, SVM, ANN, *k*-nearest neighbor (*k*-NN), and *k*-means clustering models, the random forest model performed the best, predicting childhood obesity for children under 5 years old with an accuracy of 89%.
Mfateneza et al. (2022)	• Relative to the decision tree, SVM, and logistic regression models, the random forest model performed the best, predicting infant mortality with an accuracy of 84.3%.
Mlandu et al. (2023)	• Relative to the logistic regression, decision tree, SVM, and ANN models, the random forest model performed the best, predicting dropout from the maternal, newborn, and child healthcare continuum with an accuracy of 75.7%.
Nasejje et al. (2022)	• Relative to the random survival forest (RSF) model, the deep survival neural network model (DeepSurv) model performed better in predicting child mortality before the age of 5 years old with mean C-index values of 67%–80%.
Pan et al. (2017)	• The random forest penalized logistic regression, linear discriminant analysis, and naïve Bayes models outperformed the baseline model (i.e., current paper-based risk assessment for maternal adverse births) by up to 36%. The random forest penalized logistic regression, linear discriminant analysis, and naïve Bayes models performed similarly.
Tiyyagura et al. (2022)	• The natural language processing (NLP) model identified that 3.1% of infants had injuries related to physical abuse. The NLP model yielded a positive predictive value (i.e., true positives) of 60.3% and a negative predictive value (i.e., true negatives) of 99.8% for identifying infants' injuries related to physical abuse.
	AI use to find most important variables (explanation)
Conrey et al. (2024)	• Elevated BMI and lacking strong breastfeeding intention were the most important predictors of mothers' risk of not meeting breastfeeding recommendations.

Demsash (2023)	• Late initiation of breastfeeding, mothers with no formal education, short birth intervals, poor wealth status, and unexposed to media were the top five important predictors of child mortality before the age of 5 years old.
Dey et al. (2018)	• Being poor, illiterate, and Muslim were the most important predictors of mothers' less likelihood to register pregnancies. Features most associated with lack of antenatal care coverage include being poor and illiterate, as well as lack of childbirth deliveries at health facilities.
Hazlett et al. (2023)	• Previous death of a sibling, child gender, maternal age, and malaria incidence rate in the area were the most important predictors of infant mortality.
Lee et al. (2022)	• Last norovirus detection rate, minimum temperature, and day length were the most important predictors of norovirus outbreaks.
Liu et al. (2022)	• Gestational age, mothers' abdominal circumference, mother's height, and mothers' pre-pregnancy BMI were the most important predictors for infant birth weight.
Mfateneza et al. (2022)	• Marital status, children ever born, birth order, and wealth index were the most important predictors of infant mortality.
Mlandu et al. (2023)	• Household wealth, place of residence, maternal education, and exposure to mass media were the most important predictors of dropout from the maternal, newborn, and child healthcare continuum.
Mondal et al. (2023)	• Current BMI, birth BMI, birth weight, and birth height were the most important predictors of childhood obesity.
Nasejje et al. (2022)	• Mothers' educational level, household wealth, and place of residence were the most important predictors of child mortality before the age of 5 years old.
Nti & Owusu-Boadu (2022)	• Mothers' blood sugar, age, and body temperature were the most important predictors of maternal risk of mortality during pregnancy.
Pan et al. (2017)	• Mothers' multifetal pregnancies, low pre-pregnancy weight, previous preterm birth, and age of 40 years or above were the most important predictors of mothers' experiences of adverse births.
Silva et al. (2021)	• Length of hospital stay was the most important predictor of mothers' meeting exclusive breastfeeding recommendations across different time points (i.e., at hospital discharge, when child was 3 months of age, when child was 6 months of age).
Syed et al. (2022)	• Suspected child maltreatment and suspected maternal intimate partner violence (IPV) were the most important predictors of child maltreatment or maternal IPV.
Tesfaye et al. (2019)	• Television ownership, first antenatal care, birth order, and contraceptive use were the most important predictors of mothers' skilled delivery service utilization.
Waynforth (2020)	• Birthweight, lower socioeconomic status, young maternal age, and longer duration or labor were the most important predictors of infant breathing distress.
AI use to assess treatment effects (e.g., causal inference)	
Peet et al. (2023)	• WIC reduced the risk of preterm birth, low birth weight, small size for gestational age, and infant mortality for some of the highest risk groups (i.e., risk level for poor infant health outcomes was 60%–100%).
Peet et al. (2024)	• Doula services reduced the risk of preterm birth, low birth weight, and infant mortality for the highest risk groups (i.e., risk level for poor infant health outcomes was 80%–100%).

Note. AdaBoost = adaptive boosting; ANN = artificial neural network; LASSO = least absolute shrinkage and selection operator; MNCH = maternal, newborn, and child healthcare; WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.