Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Kaitlin E. Schrote	April 14, 2025

A Bayesian Hierarchical Spatial Mapping Approach to Assess Gestational Diabetes Mellitus Risk Among Immigrant Populations in Georgia

By

Kaitlin E. Schrote

Master of Public Health

Department of Biostatistics and Bioinformatics

Emily N. Peterson, Ph.D.

(Committee Chair)

Howard H. Chang, Ph.D.

(Committee Member)

A Bayesian Hierarchical Spatial Mapping Approach to Assess Gestational Diabetes Mellitus Risk Among Immigrant Populations in Georgia

By

Kaitlin E. Schrote

M.S., Cornell University, 2013

B.A., Notre Dame of Maryland University, 2010

Committee Chair: Emily N. Peterson, Ph.D.

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health

in the Department of Biostatistics and Bioinformatics

2025

Abstract

A Bayesian Hierarchical Spatial Mapping Approach to Assess Gestational Diabetes Mellitus Risk Among Immigrant Populations in Georgia

By Kaitlin E. Schrote

Gestational diabetes mellitus (GDM) poses significant health risks for both mothers and infants, making monitoring small-area GDM prevalence critical for informing targeted public health interventions. Foreign-born women, despite generally healthier pregnancy profiles, experience higher GDM risk than US-born women. In Georgia, nearly 1 in 4 immigrants face barriers to comprehensive prenatal care and the burden of GDM is likely under-counted. To estimate the true risk of GDM, we developed a Bayesian Hierarchical Immigrant GDM (BHIG) estimation model that accounts for spatial differences and corrects for measurement error. By borrowing strength across counties and adjusting for data sparsity, the BHIG model produces robust county-level estimates of GDM risk among immigrant mothers. Model results suggests that the burden of GDM among immigrant populations in Georgia is higher than currently reported - 12.4% compared to 7.9% according to Georgia Department of Public Health records. Spatial trends reveal clusters of elevated risk and underdiagnosis in central Georgia and the Atlanta metropolitan area. This work demonstrates the importance of integrating spatial and measurement error models to better address maternal health equity.

A Bayesian Hierarchical Spatial Mapping Approach to Assess Gestational Diabetes Mellitus Risk Among Immigrant Populations in Georgia

By

Kaitlin E. Schrote

M.S., Cornell University, 2013

B.A., Notre Dame of Maryland University, 2010

Committee Chair: Emily N. Peterson, Ph.D.

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health

in the Department of Biostatistics and Bioinformatics

2025

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF TABLES

LIST OF FIGURES

1. Introduction

Amidst the maternal health crisis in Georgia, there is a critical need to comprehensively and accurately assess the burden of gestational diabetes mellitus (GDM) across the immigrant population to aid targeted interventions and identify high need populations. Research on GDM among immigrants has highlighted significant disparities and complex interactions between socioeconomic, geographic, and racial factors. Immigrant women, who generally have healthier pregnancy outcomes - such as lower rates of preterm birth and low birth weight - compared to US-born women, surprisingly face a higher risk of developing GDM (OR: 1.60 compared with US-born)(Huang et al., 2024; Adegoke et al., 2022; Shah et al., 2021). However, current approaches to assessing the risk of GDM among immigrant populations face several limitations, including (1) large geographic variations in GDM risk among immigrant mothers, (2) data limitations with small population sizes and significant uncertainty, (3) multilevel socioeconomic determinants affecting maternal outcomes, and (4) biases in reporting. GDM affects between 6% and 20% of all pregnancies in the United States (US) and is associated with significant maternal and neonatal health risks(Sperling et al., 2023; Deputy, 2018; Bower et al., 2019; Kim et al., 2013). Complications related to GDM include increased risk of cesarean section, preterm birth, eclampsia, and associated maternal and infant birth trauma (Shah et al., 2021; Kim et al., 2013; Venkatesh et al., 2022). Furthermore, GDM can have long-lasting effects on a child's health, contributing to childhood obesity, metabolic syndrome, and developmental delays (Mitanchez et al., 2015; Weintrob et al., 1996). Traditional large-scale studies often overlook geographic heterogeneity among the immigrant population and severely undercount GDM cases, leading to underestimated disparities and an incomplete risk assessment for this hard to capture population. Critical to the comprehensive assessment of immigrant maternal health disparities, small area estimation (SAE) of GDM risk among immigrant mothers at the county level can provide a more precise understanding of geographic disparities and high-risk communities, facilitating targeted public health interventions (Wakefield, 2007; Waller and Gotway, 2004; Kramer and Williamson, 2013; Waller et al., 1997). In this study, we leverage SAE methods and hierarchical measurement error modeling techniques,

to produce more reliable, localized estimates of GDM prevalence correcting for data limitations.

While studies have consistently shown that immigrant women exhibit higher risk of GDM compared to US-born women, existing studies often lack comprehensive analyses that account for the intersectionality of nativity, race/ethnicity, socioeconomic, and geographic factors, signaling a need for more nuanced investigations to address these disparities effectively (Erbetta et al., 2023; Pu et al., 2015). The US immigrant population, which accounted for 13.8% of the total US population in 2020, often faces barriers to accessing comprehensive prenatal care (Kim et al., 2013; Swartz et al., 2017). Medicaid, the largest payer for obstetric care services, covers nearly 50% of all births (Markus et al., 2013). Throughout the US, standard Medicaid provides coverage for all pregnancy-related care, encompassing the antenatal period, childbirth, and postpartum. However, under federal law, authorized immigrants in their first 5 years in the US and unauthorized immigrants are ineligible for standard Medicaid coverage, limiting them to Emergency Medicaid, a federal safety net program which excludes prenatal and postpartum care, covering only obstetric admissions(Swartz et al., 2017; DuBard and Massing, 2007; Rodriguez et al., 2020). States may choose to use their own funds to provide additional health services for immigrants, but coverage remains inconsistent (Swartz et al., 2017). Georgia, with a diverse immigrant population comprising 10% of the state's population, faces a significant challenge (U.S. Census Bureau, 2023). Over a third of these immigrants are unauthorized, and the state has one of the highest rates of uninsured individuals in the US (Migration Policy Institute (MPI), 2024a,b; Terlizzi and Cohen, 2022). Although a policy change in January 2023 allows low income authorized pregnant immigrants to qualify for standard Medicaid without a waiting period, undocumented pregnant Georgians remain restricted to Emergency Medicaid only, leading to inadequate prenatal care and an undercount of GDM cases across the state (Migration Policy Institute (MPI), 2024).

Previous studies assessing GDM risk and other maternal morbidities among immigrant populations have employed a variety of methodological approaches, including large-scale epidemio-

logical surveys, administrative health data analysis, and logistic regression models to estimate risk differentials (Janevic et al., 2014; Pu et al., 2015; Bolduc et al., 2024; Go et al., 2024). Some studies have reported a "protective effect" of immigrant status, suggesting lower rates of adverse pregnancy outcomes compared to US-born counterparts, often attributed to the "healthy immigrant effect" (Acevedo-Garcia et al., 2010; Gagnon et al., 2009). However, these findings are inconsistent, with more recent studies highlighting an elevated risk of GDM among immigrants, particularly those from Latin America, Southeast Asia, and East Asia (Shah et al., 2022a; Kim et al., 2013; Shah et al., 2022b). Methodologically, most prior work has relied on hospital discharge data or birth certificates, which are limited by underreporting, misclassification, and sometimes a lack of granular geographic detail (Backes et al., 2020). Many studies have leveraged small area estimation (SAE) techniques to capture geographic disparities in maternal health outcomes, but typically focus on broader populations without explicitly addressing immigrant-specific data limitations (Waller and Gotway, 2004; Kramer and Williamson, 2013; Yang et al., 2024; Stanhope et al., 2024). A recent study by Sun et al. (2023) applied Bayesian spatial mapping models to examine GDM incidence across Florida, identifying spatial clusters of increased risk and underscoring the value of hierarchical modeling in addressing geographic disparities. However, this study did not account for systematic reporting biases that may differentially affect immigrant populations. This gap underscores the need for refined methodological approaches that account for both geographic heterogeneity and systematic underestimation of GDM prevalence. By integrating Bayesian hierarchical modeling with measurement error correction, our study seeks to overcome these limitations and provide more reliable, localized estimates of GDM burden among immigrant populations in Georgia.

In this study, we aim to obtain small area estimates of the true risk of GDM among immigrant communities across all 159 counties in Georgia, providing a targeted understanding of the burden of GDM in this population. Our approach fuses Bayesian hierarchical disease mapping and measurement error methodologies to not only enhance the precision and robustness of GDM estimates

in the presence of error-prone data, but also to offer actionable insights for public health interventions aimed at mitigating the disproportionate burden of GDM. We use a standard Bayesian disease mapping model which allows for borrowing of strength of information across small areas and thereby reducing high degrees of uncertainty associated with smaller population sizes (Waller and Gotway, 2004; Blangiardo et al., 2020; Lawson, 2013; Wakefield, 2007; Knorr-Held, 2000; Besag et al., 1991; Riebler et al., 2016). Reported GDM counts among immigrants are systemically underestimated due to lower prenatal care utilization, leading to biased estimates of GDM risk in this population (Swartz et al., 2019; Goldfarb et al., 2017). However, we identified six gold standard (GS) counties that exhibited higher crude prevalence rates and lower variance, suggesting a sufficiently large immigrant population where we could be more confident that the sensitivity of GDM detection was fairly high compared to others, although undercounting was likely still present. This discrepancy is attributed to increased access to prenatal care services, such as mobile clinics or expanded coverage programs in these counties, which facilitate greater screening for GDM and provide more accurate and representative estimates of its prevalence among immigrant populations (Swartz et al., 2019). We assume that all reported GDM counts in non-GS counties are subject to error and use a Bayesian hierarchical disease mapping approach that accounts for measurement error across all 159 counties. This allows us to estimate county-level data quality measures, specifically sensitivity—defined as the ratio of error-prone to true GDM risk—assumed within a measurement error framework (Carroll et al., 2006; Zhang et al., 2021). We further assess the relationship between true county-level estimates of GDM prevalence and driving socio-economic and geographic determinants of health as assessed through a latent model structure which models the probability of true GDM risk as a function of key informative covariates consisting of race/ethnicity, adequate prenatal care, and rurality(Erbetta et al., 2023; Shah et al., 2021; Pu et al., 2015; Venkatesh et al., 2024; Go et al., 2024; Newman et al., 2022).

Our work makes several contributions. Firstly, our methodology introduces a novel approach to correcting reporting bias in cases where small areas have error-prone data, resulting in statistical

challenges in the development of comparative data quality metrics. By explicitly accounting for measurement (undercounting) error, our framework not only enhances the accuracy and reliability of small area estimates but also establishes a scalable and adaptable model that can be applied to other geographic regions and hard-to-capture populations facing similar data limitations. Secondly, we provide comprehensive and granular estimates of GDM risk for immigrant communities in Georgia through improving GDM surveillance to refine public health strategies targeting high-risk, underrepresented immigrant populations.

This paper is organized as follows: Section 2.1 describes the data used to obtain GDM prevalence estimates. Section 2.3 describes the GS and error-prone data used to obtain GDM estimates. Section 2.3 describes the data models assumed for the GS and error-prone GDM counts. Section 2.4 describes the model assumptions of the underlying process with incorporated sensitivity correction. Section 2.5 describes the measurement error model approach. Section 2.6 summarizes the process to obtain corrected GDM estimates and associated uncertainties. Lastly, Section 4 presents results across high and low sensitivity counties.

2. Methods

2.1. Data

Cross-tabulated vital statistics data on births - stratified by county and maternal characteristics, for Georgia residents from 2018 to 2023 were obtained from the Georgia Department of Public Health (GADPH) (Georgia Department of Public Health, Office of Health Indicators for Planning (OHIP), 2023). This dataset includes demographic, health, and prenatal care information for US-born and immigrant mothers, along with GDM diagnoses. Key variables include maternal age, race/ethnicity, nativity, insurance status, prenatal care utilization, and urban/rural residency. Exploratory findings, summarized in Table1, highlight critical demographic and health disparities that motivate the modeling assumptions used in this study. Differences in maternal age, race/ethnicity composition, and prenatal care access between US-born and immigrant mothers underscore

the need to account for socio-economic and demographic differences to better understand GDM disparities across Georgia's 159 counties.

Figure 1 illustrates the aggregated crude prevalence of GDM among immigrants across counties in Georgia from 2018-2023. Larger population counties, such as those in Metro Atlanta, generally show a GDM prevalence under 10%. Given that Metro Atlanta has a larger healthcare infrastructure and the expectation that immigrant populations there have greater access to prenatal care, higher GDM prevalence rates might be anticipated. However, this is not reflected in the crude prevalence estimates, suggesting that underdiagnosis or disparities in screening may still be present. Six counties consisting of Candler, Decatur, Forsyth, Grady, Hart, and Tift counties (outlined in red) show higher degrees of crude prevalence. These select counties have unique characteristics that likely contribute to improved GDM detection, including: (1) mobile clinics providing care to a large proportion of the immigrant population, (2) annual free clinics in southwest Georgia sponsored by Emory University Schools of Nursing and Medicine, and (3) immigrants with high socioeconomic status. These factors may lead to more comprehensive screening in these counties compared to others. As such, we refer to these as gold standard (GS) counties to which we assess potential underestimation in other areas with less access to prenatal care services for the immigrant population.

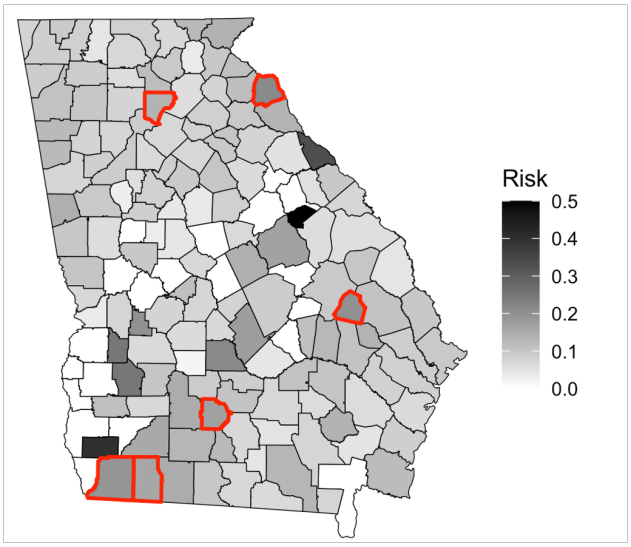

Figure 1: Mapped crude prevalence of gestational diabetes (GDM) among immigrants in Georgia, 2018-2023. Gold standard counties (Candler, Decatur, Forsyth, Grady, Hart, and Tift) are outlined in red.

Variable	US-Born	Immigrant	p-value
N	601,845	143,888	
Age			
$<$ 18 years	9,221 (1.5%)	1,090 (0.8%)	$<$ **0.001**
18 – 24 years	163,021 (27.1%)	21,863 (15.2%)	
25 – 34 years	339,256 (56.4%)	81,247 (56.5%)	
$>$ 35 years	90,347 (15.0%)	39,688 (27.6%)	
Race/Ethnicity			
Non-Hispanic White	306,887 (51.3%)	14,911 (10.4%)	$<$ **0.001**
Non-Hispanic Black	228,101 (38.1%)	24,646 (17.2%)	
Non-Hispanic Asian/Pacific Islander	5,588 (0.9%)	27,490 (19.2%)	
Non-Hispanic Other	15,094 (2.5%)	2,480 (1.7%)	
Hispanic	42,935 (7.2%)	73,444 (51.4%)	
Insurance at Delivery			
Private	298,306 (49.6%)	45,836 (31.9%)	$<$ **0.001**
Medicaid	249,872 (41.6%)	53,625 (37.3%)	
Self-Pay	12,591 (2.1%)	36,462 (25.4%)	
Other	40,207 (6.7%)	7,740 (5.4%)	
County Residency			
Urban	464,767 (77.5%)	133,037 (92.5%)	$<$ **0.001**
Rural	137,078 (22.5%)	10,851 (7.5%)	
Mean Number of Prenatal Visits	11.30 (4.81)	10.48 (4.91)	$<$ **0.001**
Prenatal Care			
Inadequate Prenatal Care	98,598 (16.4%)	34,098 (23.7%)	$<$ **0.001**
Late or No Prenatal Care	43,824 (7.5%)	17,571 (12.6%)	$<$ **0.001**
Less than 5 Prenatal Visits	41,766 (7.0%)	15,896 (11.2%)	$<$ **0.001**
Pregnancy Risk Factors			
Pre-pregnancy Diabetes	6,523 (1.1%)	1,675 (1.2%)	**0.009**
Gestational Diabetes	33,942 (5.6%)	11,316 (7.9%)	$<$ **0.001**
Pre-pregnancy Hypertension	18,997 (3.2%)	1,731 (1.2%)	$<$ **0.001**
Gestational Hypertension	55,385 (9.2%)	6,722 (4.7%)	$<$ **0.001**
Eclampsia	1,218 (0.2%)	193 (0.1%)	$<$ **0.001**

Table 1: Births by Nativity in Georgia, 2018-2023

2.2. Summary of model approach

A Bayesian hierarchical immigrant gestational diabetes (BHIG) model framework was employed to estimate county level GDM prevalence among immigrant mothers for 159 counties in Georgia. The complete model structure including full posterior conditional distributions are given in Appendix B. The model includes four key components:

1. The *data model* consists of modeling observed county specific gestational diabetes mellitus aggregated counts for 2018-2023 as binomial outcomes. For counties with error-prone data, observed cases, y_i, were modeled as the product of latent, true prevalence, θ_i and sensitivity, λ_i. For counties with gold standard data, observed counts, z_j, were modeled directly using θ_j. Further details are provided in Section 2.3.
2. The *process model* characterizes the latent prevalence of GDM by incorporating both systematic geographic variation and covariate effects. Additional details are provided in Section 2.4.
3. The *measurement error model* produces county-specific sensitivity estimates (λ_i) and improves inference. Additional details are provided in Section 2.5.
4. Estimates of corrected latent GDM counts are derived using posterior predictive distribution estimates of θ_i. This is further detailed in Section 2.6.

Figure 2 shows a graphical representation of the BHIG model set-up; refer to Appendix A for a summary of notation used.

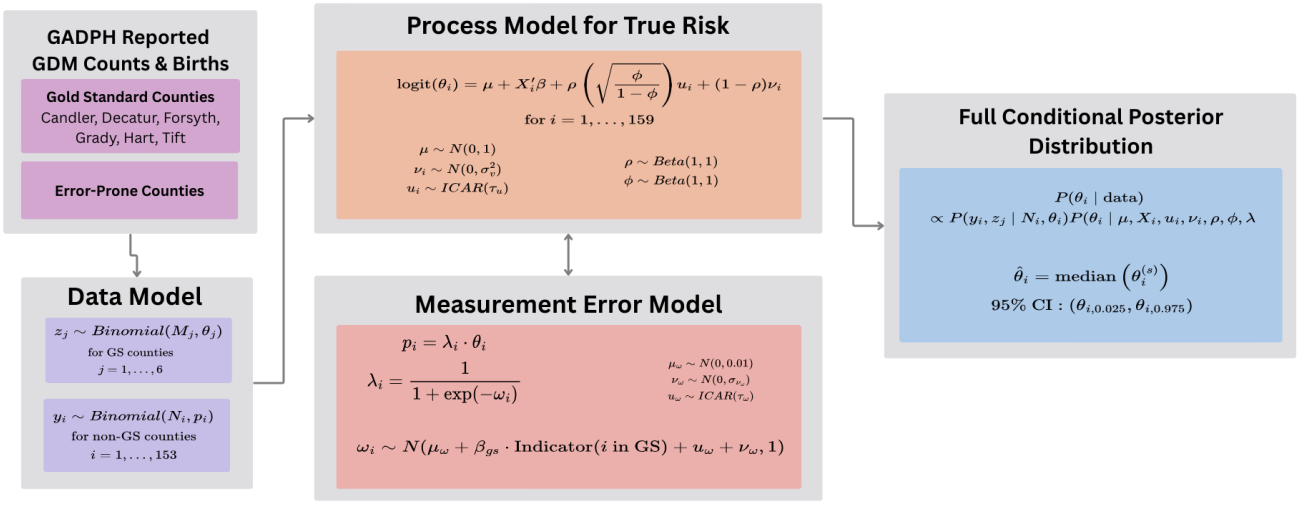

Figure 2: Bayesian hierarchical immigrant gestational diabetes mellitus (BHIG) estimation model

2.3. Data Model for Estimating Gestational Diabetes Mellitus Counts Across All Counties

The data model defines the likelihood functions that relate the observed county-level GDM counts to our parameters of interest. We distinguish between two groups of counties based on data quality. Gold standard (GS) counties (Candler, Decatur, Forsyth, Grady, Hart, and Tift) are believed to have more reliable GDM reporting systems due to mobile health units, free clinics, and socioeconomic factors. For these six counties, we assume that observed GDM counts are accurate and directly reflect the true underlying prevalence. Thus, the number of observed GDM cases, z_j is modeled as:

$$z_j \sim \text{Binomial}(M_j, \theta_j) \quad \in \text{GS County } j = 1, \ldots, 6 \tag{1}$$

where M_j is the number of live births among immigrant mothers in GS county j, and θ_j is the true prevalence of GDM.

In non-gold standard (non-GS) counties, GDM counts are assumed to be underreported. In each of these counties, the observed count of immigrant GDM cases, denoted y_i, is modeled using a Binomial distribution, where N_i is the total number of live births to immigrant mothers and p_i is the error-prone estimated prevalence of GDM in county i:

$$y_i \sim \text{Binomial}(N_i, p_i) \quad \in \text{non-GS County } i = 1, \ldots, 153 \tag{2}$$

The error-prone prevalence, p_i, is modeled as the product of county-specific sensitivity estimate λ_i and the true underlying prevalence θ_i:

$$p_i = \lambda_i \cdot \theta_i \tag{3}$$

Sensitivity measures quantify the extent of undercounting in error-prone data. This approach ensures that posterior estimates of true GDM prevalence θ_i are adjusted for measurement error, improving the robustness and interpretability of small-area estimates. This dual-likelihood framework allows us to utilize high-quality data from GS counties while accounting for potential misclassification in the rest of the state, improving the validity of county-level risk estimates across Georgia.

2.4. Process Model for Unobserved Latent Gestational Diabetes Mellitus Logit-Probabilities

We model the latent logit probability, $\text{logit}(\theta_i)$, incorporating both spatially and temporally structured and unstructured random effects shown in Eq 4. To incorporate spatial terms in our model, we consider the Reparamertized Besag-York-Mollié (BYM2) Model, which allows us to estimate the prevalence of GDM, weighting trends in the neighboring counties (Riebler, A. et al., 2016; Besag et al., 1991; Knorr-Held, 2000; Waller et al., 1997). We denote v_i to represent a spatially unstructured random effect term that is independent, identically, and normally distributed centered around zero, $v_i \sim N(0, \sigma_v^2)$. The spatially structured term, denoted u_i is modeled assuming an intrinsic conditional autoregressive (ICAR) prior, which assumes complete correlation between neighboring areas. The spatial covariance matrix, W is written as a function of an N x N adjacency matrix where entries i,i are zero and the off-diagonal elements are 1 if counties i and j are neighbors and 0 otherwise. D is the N x N diagonal matrix where entries i,i are the number of neighbors of county i and the off-diagonal entries are 0. Lastly, τ_u denotes the smoothing parameter. The parameter ρ is a mixing parameter that controls the contribution of the spatially structured random

effect u_i and the unstructured random effect v_i in the model. Specifically, ρ balances the amount of variance attributed to spatially structured effects versus unstructured effects, with higher values of ρ indicating a greater reliance on spatial structure. The parameter ϕ controls the proportion of variance attributed to the spatially structured random effect u_i by influencing the relative contribution of spatial and unstructured noise to the total model variance.

$$\begin{aligned}
&\text{logit}(\theta_i) = \mu + X_i'\beta + \rho\left(\sqrt{\frac{\phi}{1-\phi}}\right)u_i + (1-\rho)v_i, \quad \text{for } i = 1,\ldots,159 \\
&\mu \sim N(0,1) \quad \text{Intercept} \\
&\rho \sim Beta(1,1) \quad \text{Structured-Unstructured Balance} \\
&\phi \sim Beta(1,1) \quad \text{Proportion of Spatial Variance} \\
&v_i \sim N(0,\sigma_v^2) \quad \text{Unstructured spatial noise} \\
&u_i \sim N\left(0,[\tau_u(D-W)]^{-1}\right) \quad \text{ICAR prior for spatial auto-correlation}
\end{aligned} \tag{4}$$

2.5. Measurement Error Model

We model λ_i on the logit-transformed scale to extend the parameter space from (0,1) to the real line. We model the logit-transformed λ_i as a spatially structured ICAR random term which allows for county-specific deviations but also allows for sharing of information across counties, such that counties closer to gold-standard counties have higher levels of sensitivity as the benefit from expanded prenatal care access in neighboring counties. The measurement error model is given in Eq. 5

$$
\begin{aligned}
\lambda_i &= \frac{1}{1+\exp(-\omega_i)} \\
\omega_i &\sim N(\mu_\omega + \beta_{gs} \cdot \text{Indicator } (i \text{ in GS}) + u_\omega + v_\omega, 1) \\
\mu_\omega &\sim N(0, 0.01) \quad \text{Intercept} \\
\beta_{gs} &\sim \text{Truncated } N(1, 1; [0, 5]) \\
v_\omega &\sim N(0, \sigma^2_{v_\omega}) \quad \text{Unstructured spatial noise} \\
u_\omega &\sim N\left(0, [\tau_\omega (D - W)]^{-1}\right) \quad \text{ICAR prior for spatial auto-correlation}
\end{aligned}
\tag{5}
$$

2.6. Derivation of Corrected GDM Counts and Associated Uncertainty

To estimate the true prevalence of GDM among immigrant populations, we obtain posterior median estimates of θ_i, which account for measurement error due to imperfect sensitivity λ_i. The full conditional posterior distribution for θ_i is given by:

$$
P(\theta_i \mid \text{data}) \propto P(y_i, z_j \mid N_i, M_j, \theta_i) P(\theta_i \mid \mu, X_i, u_i, v_i, \rho, \phi, \lambda_i), \tag{6}
$$

where $P(y_i, z_j \mid N_i, M_j, \theta_i)$ follows a Binomial likelihood:

$$
y_i \mid N_i, \theta_i \sim \text{Binomial}(N_i, \theta_i), \tag{7}
$$

and $P(\theta_i \mid \mu, \beta, u, v, \rho, \phi, \lambda_i)$ represents the prior distribution for θ_i, incorporating spatial smoothing and measurement error adjustments through λ_i.

From this full conditional posterior distribution, we obtain posterior median estimates of θ_i, denoted as:

$$
\hat{\theta}_i = \text{median}\left(\theta_i^{(s)}\right), \tag{8}
$$

where $\theta_i^{(s)}$ are posterior samples from $P(\theta_i \mid \text{data})$. These posterior median estimates provide corrected GDM rates for immigrant populations, addressing undercounting biases and improving the robustness of small-area prevalence estimation while incorporating spatial correlations across

counties.

2.7. Simulation Study: Evaluating Measurement Error Correction

To assess the ability of our BHIG model to recover the true prevalence of GDM while accounting for undercounting, we conducted a simulation study. The simulation generates synthetic county-level immigrant GDM data under varying degrees of underreporting, applying the two-stage model to determine whether the posterior estimates correctly adjust for measurement error.

Data Generation We generate simulated county-level immigrant GDM counts across $N = 159$ counties in Georgia. For ease of readability, we denote generated samples with (s) and fixed preset values with $*$. The true GDM prevalence $\theta_i^{(s)}$ in each county i follows a logistic regression model incorporating covariate information and spatial dependence:

$$\text{logit}(\theta_i^{(s)}) = \mu^* + X_i^{*\prime}\beta^* + \rho^* \left(\sqrt{\frac{\phi^*}{1-\phi^*}} \right) u_i^{(s)} + (1-\rho^*) v_i^{(s)}, \quad i = 1, \ldots, 159 \tag{9}$$

where μ is a fixed intercept term, X_i represents known county-level covariates, β^* are regression coefficients, $u_i^{(s)}$ is a spatially structured random effect (modeled using an intrinsic conditional autoregressive (ICAR) prior), and $v_i^{(s)}$ is an independent Gaussian noise term. The prevalence of GDM prone to errors $p_i^{(s)}$ is systematically lower than $\theta_i^{(s)}$, following the sensitivity relationship $p_i^{(s)} = \lambda_i^{(s)} \cdot \theta_i^{(s)}$, where $\lambda_i^{(s)}$ represents the county-specific sensitivity parameter, controlling the degree of undercounting in all counties. The error-prone counts $y_i^{(s)}$ are then simulated as:

$$y_i \sim \text{Binomial}(N_i, p_i) \tag{10}$$

Simulation Scenarios: We impose increasing degrees of measurement error across three scenarios shown below. Scenario 1 imposes a small degree of error, scenario 2 imposes a moderate degree of error, and scenario 3 imposes the highest degree of error.

$$\text{Scenario 1: } \lambda_i^{(s)} \sim Unif(0.8, 1)$$

$$\text{Scenario 2: } \lambda_i^{(s)} \sim Unif(0.5, 1)$$

$$\text{Scenario 3: } \lambda_i^{(s)} \sim Unif(0.2, 0.8)$$

For each scenario, we generate $s = 1, ..., 200$ Monte Carlo datasets and fit the full BHIG model approach outlined in Eqs 2- 8.

Evaluation Metrics The ability of the model to recover the true prevalence θ_i is assessed using summary metrics of mean error, mean absolute error, mean relative error, median error, median absolute error, median relative error, and 95% inside coverage probabilities.

3. Computation

We used GADPH-reported births for 159 counties in Georgia from years 2018-2023. For model processing and output, a Markov Chain Monte Carlo (MCMC) algorithm samples from the posterior distribution of the parameters via the software *Nimble* (de Valpine et al., 2017). Ten parallel chains were run with a total of 60,000 iterations in each chain. Of these, the first 10,000 iterations in each chain were discarded as burn-in, leaving 50,000 samples per chain. To reduce autocorrelation, we thinned the samples by retaining every 10th iteration after burn-in. Standard diagnostic checks, including traceplots were used to assess convergence (Plummer, 2017; Gelman and Rubin, 1992; Vehtari et al., 2021; Su and Yajima, 2020; Rue and Held, 2005; de Valpine et al., 2017).

4. Results

4.1. Simulation Results

The results demonstrate that the BHIG model with measurement error correction successfully recovers the true prevalence θ_i across counties, particularly under mild and moderate undercounting scenarios. Severe undercounting ($\lambda_i < 0.6$) leads to increased uncertainty, though the model still provides reasonable coverage probabilities. Across all scenarios, incorporating gold-standard data improves the precision of prevalence estimates, highlighting the importance of leveraging high-quality data sources in disease mapping.

Table 2 presents summary error metrics, including the median absolute error across undercounting scenarios, showing improved estimation accuracy as λ_i approaches 1.

Scenario	**Mean Error**	**Median Error**	**Median Abs Error**	**MSE**	**95% Inside Coverage**
Low Bias	0.005	0.015	0.019	0.022	95.20%
Moderate Bias	-0.024	-0.043	0.033	0.027	94.10%
High Bias	-0.062	-0.075	0.069	0.054	91.11%

Table 2: Summary of model performance across 100 simulated county-level datasets under varying levels of bias imposed on error-prone counties.

These findings underscore the necessity of correcting for systematic undercounting in small-area disease prevalence estimates. By integrating spatial smoothing and measurement error modeling, our Bayesian approach provides a more robust framework for understanding maternal health disparities in immigrant populations.

4.2. Global Parameter Estimates

Global parameters consist of the global variance terms σ_v^2 and σ_ω^2, the global scaling parameters ϕ and ρ, and the global intercepts μ and μ_ω. Table3 shows posterior model-estimates and credible intervals for the global parameters.

Parameters	Estimate	95% CI Lower Bound	95% CI Upper Bound
Fixed Effects (Odds Ratios)			
μ	0.151	0.111	0.210
Prenatal Care	0.850	0.350	2.128
Hispanic Ethnicity	0.987	0.524	1.982
Rurality	1.058	0.844	1.316
Sensitivity Components			
μ_ω	0.320	0.134	0.505
β_{gs}	1.214	0.081	3.024
σ^2_ω	0.011	0.001	0.243
Spatial and Variance Components			
σ^2_v	0.332	0.220	1.028
ϕ	0.244	0.008	0.863
ρ	0.177	0.006	0.742

Table 3: Global parameter estimates included in the BHIG model

4.3. Estimated GDM Sensitivity Across Georgia

The estimated sensitivity $\hat{\lambda}_i$ of GDM case detection varied across Georgia, with distinct spatial patterns emerging. As shown in Figure 3, sensitivity estimates ranged from 0.26 to 0.85, and were generally highest in the southwest region of the state, which is an expected result. Among the gold standard counties, Hart County had the highest estimated sensitivity at 0.848 (95% CI: 0.284 - 0.989), followed closely by Grady County at 0.845 (95% CI: 0.295 - 0.988), Candler County at 0.835 (95% CI: 0.267 - 0.988), Forsyth County at 0.818 (95% CI: 0.251 - 0.987), and Tift County at 0.812 (95% CI: 0.234 - 0.987). Decatur County had a slightly lower estimate with sensitivity value of 0.782 (95% CI: 0.195 - 0.0.984). In contrast, Fayette County exhibited the lowest estimated sensitivity at 0.263 (95% CI: 0.141 - 0.486), substantially lower than any of the gold standard counties. The overall spatial distribution of sensitivity estimates suggests regional differences in GDM case detection across the state. Additionally, the findings highlight the importance of localized adjustments when estimating GDM prevalence, particularly in areas with limited gold standard data.

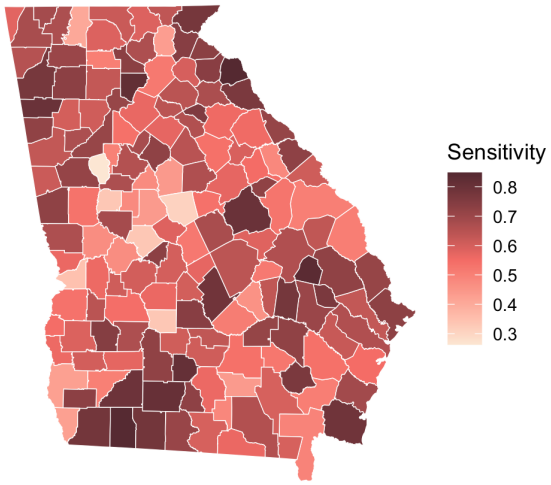

Figure 3: Mapped posterior median sensitivity estimates for gestational diabetes in Georgia, 2018-2023.

4.4. Posterior Probability and Variance Estimates

The spatial distribution of the posterior median probability estimates (Figure 4, left) indicates notable regional variation in the estimated probability of GDM. Higher probability estimates are observed in the southwestern regions of Georgia, as well as in Hart County in northeast Georgia. The posterior median variance estimates (Figure 4, right) reveal lower uncertainly in the gold standard counties, as expected. Counties with the highest variance estimates tend to correspond to areas with smaller immigrant populations and thus fewer observed births and GDM cases. Together these maps highlight the spatial heterogeneity in both the probability of GDM and the confidence in these estimates.

To further assess the impact of sensitivity, Figure 5 presents the posterior median probability estimates of GDM prevalence under two modeling approaches: without incorporation of the Measurement Error Model (left) and with this additional sensitivity adjustment (middle). The rightmost panel displays the difference between these models, highlighting areas where prevalence estimates increased the most, following the incorporation of sensitivity. In the unadjusted model (left panel), the GDM prevalence estimates exhibit spatial heterogeneity throughout Georgia, with much higher estimates observed in the north Atlanta suburbs and southwestern counties.

After incorporating sensitivity (middle panel), overall prevalence estimates increased, with more pronounced increases where sensitivity was previously lower. The difference map (right panel) indicates the magnitude of these changes, with the largest increases in estimated GDM prevalence occurring in dark blue counties, primarily south of the Atlanta metropolitan region. Counties in red, concentrated in southwestern Georgia where several gold standard counties are located, exhibit the smallest increases in estimated GDM prevalence. These results suggest that incorporating sensitivity refines spatial estimates, particularly in regions where case ascertainment is less certain.

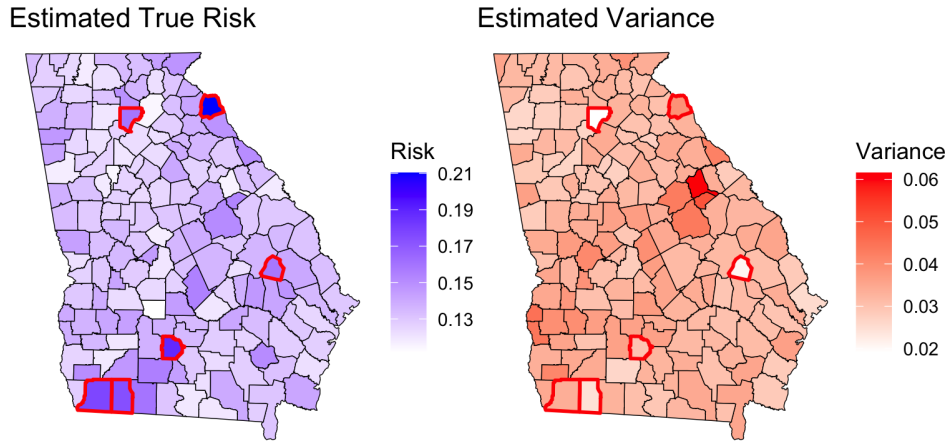

Figure 4: Mapped true posterior median probability (left) and true posterior median variance (right) estimates

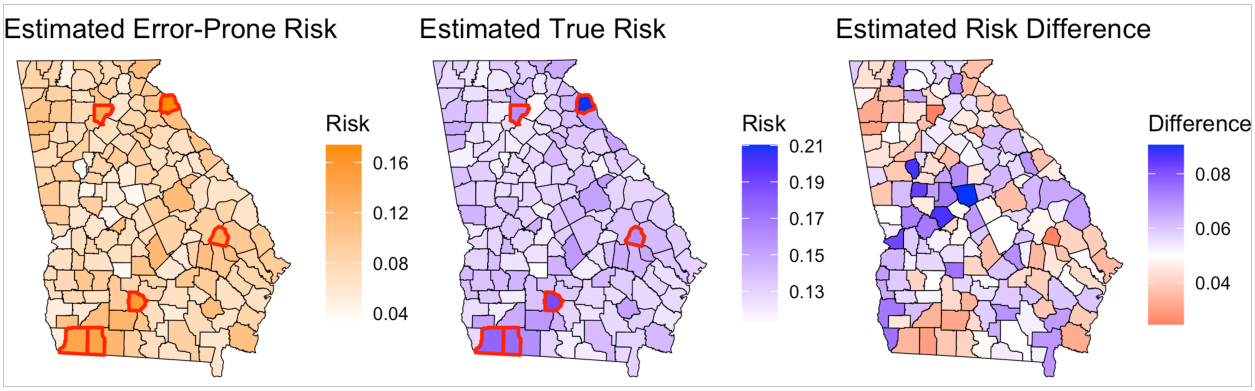

Figure 5: Mapped comparison of posterior median model estimates without and with incorporation of sensitivity. (Left) Posterior median probability estimates for the model without incorporation of sensitivity; (Middle) Posterior median probability estimates for the model with sensitivity adjustment; (Right) Difference between the two models, highlighting areas of change.

4.5. Estimated Additional GDM Cases Across Georgia

Table4 compares the GADPH-reported number of GDM cases among immigrants in Georgia from 2018 to 2023 along with estimates from the BHIG model. GADPH reported 11,316 cases of GDM during this period of time, with a state-wide risk of 7.9% among immigrants. In contrast, the BHIG model estimates a 12.4% GDM risk during the same time period, suggesting approximately 6,600 missed cases of GDM among immigrant mothers.

	Cases 2018-2023	**95% CI**	**Additional Cases**	**GDM Risk**
GADPH Reported	11,316	-	-	7.9%
BHIG Estimates	17,905	(12,079–28,908)	6,589	12.4%

Table 4: Comparison of GDM cases reported by GADPH and estimated by the BHIG model, 2018-2023

5. Discussion

This study demonstrates substantial underreporting of GDM cases among immigrant mothers in Georgia. Estimates from our BHIG model, which incorporates measurement error suggests a "true" GDM prevalence of 12.4% among immigrants, compared to the 7.9% reported in GADPH records. This discrepancy - amounting to almost 6,600 missed cases between 2018 and 2023 - confirms prior research indicating that GDM is frequently underestimated in immigrant populations due to inconsistent screening and barriers to healthcare access (Janevic et al., 2022; Swartz et al., 2017).

Geographic variation in the extent of underreporting is notable. Contrary to our expectations, urban counties, particularly those in the Metro Atlanta region, exhibited more pronounced underestimation than rural counties. While these areas typically have more robust prenatal care infrastructure, our model indicates gaps in diagnosis or documentation such that the reported GDM prevalence in the Metro Atlanta region remains below what should be expected based on existing literature (Pu et al., 2015; Shah et al., 2022a). Factors likely contributing to this including delayed initiation of prenatal care, which is commonly associated with financial, legal, and language barriers among immigrant communities (Bustamante et al., 2022; Korinek and Smith, 2011; Shi et al., 2009).

Additionally, exclusion of certain immigrant groups from Medicaid and other related programs reduces opportunities for early and consistent screening, perpetuating disparities in maternal health outcomes.

In contrast to urban counties, GADPH-reported GDM prevalence in rural counties, particularly those in Southwest Georgia, more closely aligns with estimates from the literature, and our model's estimates show only a modest increases in cases. This pattern suggests more complete GDM reporting in these areas, potentially due to higher rates of screening facilitaed by Georgia Farmworker Health Program (GFHP) clinics, several of which deploy mobile units that provide point-of-care testing for underserved populations (Georgia Department of Community Health, 2025). Because GDM is typically recorded on birth certificates based on clinical diagnoses documented in prenatal records, increased screening and diagnostic care may directly contribute to improved case ascertainment. However, despite this more comprehensive GDM reporting, these regions still face significant barriers to prenatal care access, which may contribute to worse maternal and neonatal health outcomes. One key challenge is the high proportion of Medicaid-eligible individuals who do not enroll, instead relying on self-pay for prenatal care, which greatly limits their ability access comprehensive services (Bustamante et al., 2022; Luque et al., 2018). This phenomenon is particularly common among Latina and Asian immigrant populations, where concerns about Medicaid enrollment persist due to fear surrounding immigration status (Daudi, 2020). The 2019 changes to the "public charge" rule made it more difficult for immigrants to obtain a green card or permanent residence if they used public benefits such as Medicaid or SNAP (Bernstein et al., 2020). Although these changes were reversed in 2021, fear and misinformation continue to deter many immigrants from enrolling in Medicaid (White House, 2021; Bustamante et al., 2022; Wang et al., 2022). This reluctance to seek publicly funded healthcare may lead to delays in prenatal care initiation and reduced access to GDM screening, ultimately exacerbating disparities in maternal and infant health outcomes. Addressing these barriers requires targeted outreach to immigrant communities, clear communication about Medicaid eligibility policies, and assurances that using prenatal care ser-

vices will not negatively impact immigration status.

This study has several limitations that should be considered when interpreting the findings. First, our model relies on assumptions about spatial patterns and reporting mechanisms, which may introduce biases if the true distribution of GDM cases does not align with these assumptions. For example, if reporting errors are not spatially structured or sensitivity varies in a way not captured by the model, the sensitivity estimates may be systemically biased, either over- or underestimating sensitivity in certain counties. Another key limitation is the lack of individual-level data on immigration status, healthcare utilization, and gestational diabetes screening practices. Without these data, our model primarily captures geographic trends rather than person risk factors, potentially obscuring important sociodemographic determinants of GDM underreporting and prevalence.

Future research should explore barriers to prenatal care and GDM diagnoses in both urban and rural immigrant populations to better understand disparities in screening and management. Additionally, with the recent Medicaid postpartum expansion to 12 months in Georgia, it would be valuable to investigate how many individuals diagnosed with GDM receive postpartum diabetes screening and subsequently are diagnosed with Type 2 diabetes mellitus. Estimating the proportion of GDM cases that represent undiagnosed pre-pregnancy pre-diabetes could provide critical insights into early intervention strategies for improving long-term health outcomes.

This study highlights significant underreporting of GDM cases among immigrant mothers in Georgia, underscoring the need for more accurate data collection and reporting. To better understand these disparities, we employed a Bayesian spatial mapping approach which adjusts for measurement error, which allows for the estimation of GDM prevalence while accounting for spatial dependencies and variations in data quality across counties. This method enables more precise prevalence estimates by borrowing strength from neighboring regions and incorporating sensitivity adjustments, particularly in areas where data may be error-prone. By using this approach, we can

more effectively identify high-risk areas and inform targeted interventions. Addressing gaps in screening and reporting remains essential for improving maternal health outcomes in immigrant communities and reducing the complications associated with untreated GDM.

Bibliography

X. Huang, K. Lee, M. C. Wang, N. S. Shah, A. M. Perak, K. K. Venkatesh, W. A. Grobman, S. S. Khan, Maternal Nativity and Preterm Birth, JAMA Pediatrics 178 (2024) 65–72. URL: https://doi.org/10.1001/jamapediatrics.2023.4907. doi:10.1001/jamapediatrics.2023.4907.

T. M. Adegoke, L. F. Pinder, N. Ndiwane, S. E. Parker, O. Vragovic, C. D. Yarrington, Inequities in adverse maternal and perinatal outcomes: the effect of maternal race and nativity, Maternal and Child Health Journal 26 (2022) 823–833.

N. S. Shah, M. C. Wang, P. M. Freaney, A. M. Perak, M. R. Carnethon, N. R. Kandula, E. P. Gunderson, K. M. Bullard, W. A. Grobman, M. J. O'Brien, et al., Trends in gestational diabetes at first live birth by race and ethnicity in the us, 2011-2019, Jama 326 (2021) 660–669. URL: https://doi.org/10.1001/jama.2021.7217. doi:10.1001/jama.2021.7217.

M. M. Sperling, S. A. Leonard, Y. J. Blumenfeld, S. L. Carmichael, J. Chueh, Prepregnancy body mass index and gestational diabetes mellitus across asian and pacific islander subgroups in california, AJOG Global Reports 3 (2023) 100148. URL: https://www.sciencedirect.com/science/article/pii/S266657782200096X. doi:10.1016/j.xagr.2022.100148.

N. P. Deputy, Prevalence and changes in preexisting diabetes and gestational diabetes among women who had a live birth—united states, 2012–2016, MMWR. Morbidity and mortality weekly report 67 (2018) 1201–1207.

J. K. Bower, B. N. Butler, S. Bose-Brill, J. Kue, C. L. Wassel, Racial/ethnic differences in diabetes screening and hyperglycemia among us women after gestational diabetes, Preventing Chronic Disease 16 (2019).

S. Y. Kim, W. Sappenfield, A. J. Sharma, H. G. Wilson, C. L. Bish, H. M. Salihu, L. J. England, Racial/ethnic differences in the prevalence of gestational diabetes mellitus and maternal

overweight and obesity, by nativity, florida, 2004-2007, Obesity 21 (2013) E33–E40. URL: https://doi.org/10.1002/oby.20025. doi:10.1002/oby.20025.

K. K. Venkatesh, C. D. Lynch, C. E. Powe, M. M. Costantine, S. F. Thung, S. G. Gabbe, W. A. Grobman, M. B. Landon, Risk of Adverse Pregnancy Outcomes Among Pregnant Individuals With Gestational Diabetes by Race and Ethnicity in the United States, 2014-2020, JAMA 327 (2022) 1356–1367. doi:10.1001/jama.2022.3189.

D. Mitanchez, C. Yzydorczyk, B. Siddeek, F. Boubred, M. Benahmed, U. Simeoni, The offspring of the diabetic mother – short- and long-term implications, Best Practice Research Clinical Obstetrics Gynaecology 29 (2015) 256–269. URL: https://www.sciencedirect.com/science/article/pii/S1521693414001631. doi:https://doi.org/10.1016/j.bpobgyn.2014.08.004, non-Communicable Diseases in Maternal Fetal Medicine: Volume II.

N. Weintrob, M. Karp, M. Hod, Short- and long-range complications in offspring of diabetic mothers, Journal of Diabetes and its Complications 10 (1996) 294–301. URL: https://www.sciencedirect.com/science/article/pii/1056872795000801. doi:https://doi.org/10.1016/1056-8727(95)00080-1, from Insulin Resistance to Antidiabetic Efficacy of Benfluorex.

J. Wakefield, Disease mapping and spatial regression with count data, Biostatistics 8 (2007) 158–183. doi:10.1093/biostatistics/kx1008.

L. A. Waller, C. A. Gotway, Applied Spatial Statistics for Public Health Data, John Wiley & Sons, Inc., 2004. doi:10.1002/0471662682.

M. R. Kramer, R. Williamson, Multivariate bayesian spatial model of preterm birth and cardiovascular disease among georgia women: Evidence for life course social determinants of health, Spatial and Spatio-temporal Epidemiology 6 (2013) 25–35. URL: https://www.sciencedirect.com/science/article/pii/S1877584513000294. doi:10.1016/j.sste.2013.05.002.

L. A. Waller, B. P. Carlin, H. Xia, A. E. Gelfand, Hierarchical spatio-temporal mapping of diease rates, Journal of the American Statistical Association 92 (1997) 607–617.

K. Erbetta, J. Almeida, K. A. Thomas, Racial/Ethnic and Nativity Inequalities in Gestational Diabetes Mellitus: The Role of Psychosocial Stressors, Women's Health Issues 33 (2023) 600–609. URL: https://doi.org/10.1016/j.whi.2023.06.007. doi:10.1016/j.whi.2023.06.007.

J. Pu, B. Zhao, E. J. Wang, V. Nimbal, S. Osmundson, L. Kunz, R. A. Popat, S. Chung, L. P. Palaniappan, Racial/ethnic differences in gestational diabetes prevalence and contribution of common risk factors, Paediatric and Perinatal Epidemiology 29 (2015) 436–443. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ppe.12209. doi:10.1111/ppe.12209. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/ppe.12209.

J. J. Swartz, J. Hainmueller, D. Lawrence, M. I. Rodriguez, Expanding prenatal care to unauthorized immigrant women and the effects on infant health, Obstetrics & Gynecology 130 (2017) 938–945. URL: https://journals.lww.com/greenjournal/fulltext/2017/11000/expanding_prenatal_care_to_unauthorized_immigrant.2.aspx.

A. R. Markus, E. Andres, K. D. West, N. Garro, C. Pellegrini, Medicaid covered births, 2008 through 2010, in the context of the implementation of health reform, Women's Health Issues 23 (2013) e273–e280.

C. A. DuBard, M. W. Massing, Trends in emergency medicaid expenditures for recent and undocumented immigrants, Jama 297 (2007) 1085–1092. URL: https://doi.org/10.1001/jama.297.10.1085. doi:10.1001/jama.297.10.1085.

M. I. Rodriguez, J. J. Swartz, D. Lawrence, A. B. Caughey, Extending delivery coverage to include prenatal care for low-income, immigrant women is a cost-effective strategy, Women's Health Issues 30 (2020) 240–247. URL: https://doi.org/10.1016/j.whi.2020.02.004. doi:10.1016/j.whi.2020.02.004.

U.S. Census Bureau, ACS 1-Year Estimates: Selected Characteristics of the Native and Foreign-Born Populations in the United States, 2023. URL: https://data.census.gov/table/ACSST1Y2023.S0501?q=Georgia, accessed 3/21/2024.

Migration Policy Institute (MPI), Georgia Immigration Data Profile, 2024a. URL: https://www.migrationpolicy.org/data/state-profiles/state/income/GA, accessed: 3-21-2024.

Migration Policy Institute (MPI), Profile of the Unauthorized Population: Georgia, 2024b. URL: https://www.migrationpolicy.org/data/unauthorized-immigrant-population/state/GA, accessed: 3-21-2024.

E. P. Terlizzi, R. A. Cohen, NHSR 176: Geographic Variation in Health Insurance Coverage: United States, 2021 (2022).

Migration Policy Institute (MPI), Medical assistance programs for immigrants in various states, 2024. URL: https://www.migrationpolicy.org/data/unauthorized-immigrant-population/state/GA, accessed: 7-2-2024.

T. Janevic, L. Borrell, D. Savitz, S. E. Echeverria, A. Rundle, Ethnic enclaves and gestational diabetes among immigrant women in new york city, Social Science & Medicine 120 (2014) 180–189.

M. L. F. Bolduc, C. I. Mercado, Y. Zhang, E. A. Lundeen, N. D. Ford, K. M. Bullard, D. C. Carty, Gestational Diabetes Prevalence Estimates from Three Data Sources, 2018, Maternal and Child Health Journal (2024). URL: https://doi.org/10.1007/s10995-024-03935-1. doi:10.1007/s10995-024-03935-1.

M. Go, N. Sokol, L. Ward, M. Anderson, S. Sun, Characterizing sociodemographic disparities and predictors of gestational diabetes mellitus among asian and native hawaiian or other pacific islander pregnant people: an analysis of prams data, 2016–2022, BMC Pregnancy and Childbirth 24 (2024) 833.

D. Acevedo-Garcia, L. M. Bates, T. L. Osypuk, N. McArdle, The effect of immigrant generation and duration on self-rated health among us adults 2003–2007, Social science & medicine 71 (2010) 1161–1172.

A. J. Gagnon, M. Zimbeck, J. Zeitlin, R. Collaboration, et al., Migration to western industrialised countries and perinatal health: a systematic review, Social science & medicine 69 (2009) 934–946.

L. M. Shah, Y. A. Kwapong, E. Boakye, S. M. Ogunwole, W. L. Bennett, R. S. Blumenthal, A. G. Hays, M. J. Blaha, K. Nasir, S. Zakaria, et al., Racial disparities in prevalence of gestational diabetes by maternal nativity and length of us residence in an urban low-income population in the united states, CJC open 4 (2022a) 540–550.

N. S. Shah, M. C. Wang, N. R. Kandula, M. R. Carnethon, E. P. Gunderson, W. A. Grobman, S. S. Khan, Gestational diabetes and hypertensive disorders of pregnancy by maternal birthplace, American journal of preventive medicine 62 (2022b) e223–e231.

E. P. Backes, S. C. Scrimshaw, E. National Academies of Sciences, Medicine, et al., Issues in measuring outcomes by birth settings: Data and methods, in: Birth Settings in America: Outcomes, Quality, Access, and Choice, National Academies Press (US), 2020.

N. Yang, H. S. Quick, S. J. Melly, A. M. Mullin, Y. Zhao, J. Edwards, J. E. Clougherty, L. H. Schinasi, H. H. Burris, Spatial patterning of spontaneous and medically indicated preterm birth in philadelphia, American journal of epidemiology 193 (2024) 469–478.

K. K. Stanhope, T. Stallworth, A. D. Forrest, D. Vuncannon, G. Juarez, S. L. Boulet, F. Geary, A. L. Dunlop, S. C. Blake, V. L. Green, D. J. Jamieson, Planning for the forgotten fourth trimester of pregnancy: A parallel group randomized control trial to test a postpartum planning intervention vs. standard prenatal care, Contemporary Clinical Trials 143 (2024) 107586. URL: https://www.sciencedirect.com/science/article/pii/S1551714424001691. doi:10.1016/j.cct.2024.107586.

N. Sun, Z. Bursac, I. Dryden, R. Lucchini, S. Dabo-Niang, B. Ibrahimou, Bayesian spatiotemporal modelling for disease mapping: an application to preeclampsia and gestational diabetes in Florida, United States, Environmental Science and Pollution Research 30 (2023) 109283–109298. URL: https://doi.org/10.1007/s11356-023-29953-0. doi:10.1007/s11356-023-29953-0.

M. Blangiardo, M. Cameletti, M. Pirani, G. Corsetti, M. Battaglini, G. Baio, Estimating weekly excess mortality at sub-national level in italy during the covid-19 pandemic, PLOS ONE 15 (2020) 1–15. URL: https://doi.org/10.1371/journal.pone.0240286. doi:10.1371/journal.pone.0240286.

A. B. Lawson, Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology, 2nd ed., CRC Press Tayloy & Francis Group, 2013.

L. Knorr-Held, Bayesian modelling of inseparable space-time variation in disease risk, Statistics in Medicine 19 (2000) 2555–2567. URL: http://onlinelibrary.wiley.com/doi/abs/10.1002/1097-0258%2820000915/30%2919%3A17/18%3C2555%3A%3AAID-SIM587%3E3.0.CO%3B2-%23. doi:10.1002/1097-0258(20000915/30)19:17/18<2555::AID-SIM587>3.0.CO;2-#, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1097-0258%2820000915/30%2919%3A17/18%3C2555%3A%3AAID-SIM587%3E3.0.CO%3B2-%23.

J. Besag, J. York, A. Mollié, Bayesian image restoration, with two applications in spatial statistics, Annals of the institute of statistical mathematics 43 (1991) 1–20.

A. Riebler, S. H. Sørbye, D. Simpson, H. Rue, An intuitive bayesian spatial model for disease mapping that accounts for scaling, Statistical methods in medical research 25 (2016) 1145–1165.

J. J. Swartz, J. Hainmueller, D. Lawrence, M. I. Rodriguez, Oregon's expansion of prenatal care

improved utilization among immigrant women, Maternal and child health journal 23 (2019) 173–182.

S. S. Goldfarb, W. Smith, A. E. Epstein, S. Burrows, M. Wingate, Disparities in prenatal care utilization among us versus foreign-born women with chronic conditions, Journal of Immigrant and Minority Health 19 (2017) 1263–1270.

R. J. Carroll, D. Ruppert, L. A. Stefanski, C. M. Crainiceanu, Measurement Error in Nonlinear Models, second ed., Chapman & Hall/CRC, 2006.

K. Zhang, J. Liu, Y. Liu, P. Zhang, R. J. Carroll, Bayesian adjustment for measurement error in an offset variable in a Poisson regression model, Statistical Modelling (2021) 1–18. URL: https://doi.org/10.1177/1471082X211008011.

K. K. Venkatesh, X. Huang, N. A. Cameron, L. C. Petito, J. Joseph, M. B. Landon, W. A. Grobman, S. S. Khan, Rural–urban disparities in pregestational and gestational diabetes in pregnancy: Serial, cross-sectional analysis of over 12 million pregnancies, BJOG: An International Journal of Obstetrics & Gynaecology 131 (2024) 26–35.

S. D. Newman, C. Bui, L. J. Bradley, M. Hernandez-Reif, The Effects of Race and Socioeconomics on the Relationship Between Gestational Diabetes Mellitus and Birth Outcomes: An Analysis of Southern US PRAMS data, Maternal and Child Health Journal 26 (2022) 2476–2484. URL: https://doi.org/10.1007/s10995-022-03513-3. doi:10.1007/s10995-022-03513-3.

Georgia Department of Public Health, Office of Health Indicators for Planning (OHIP), Vital Records Birth Certificates, 2018-2023. Accessed 5/30/2024.

Riebler, A., Sørbye, S.H., Simpson, D., Rue, H., An Intuitive Bayesian Spatial Model for Disease Mapping That Accounts for Scaling, Statistical Methods in Medical Research 25 (2016) 1145–65.

L. A. Waller, B. P. Carlin, H. Xia, A. E. Gelfand, Hierarchical spatio-temporal mapping of disease rates, Journal of the American Statistical Association 92 (1997) 607–617.

P. de Valpine, D. Turek, V. Poynor, J. Franklin, NIMBLE: A Modular Framework for Building and Applying MCMC Methods in R, The R Journal 9 (2017) 505–518.

M. Plummer, JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling (2017).

A. Gelman, D. B. Rubin, Inference from Iterative Simulation Using Multiple Sequences, Statistical Science 7 (1992) 457–472. URL: http://www.jstor.org/stable/2246093, publisher: Institute of Mathematical Statistics.

A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, P.-C. Bürkner, Rank-normalization, folding, and localization: An improved r for assessing convergence of mcmc (with discussion), Bayesian analysis 16 (2021) 667–718.

Y.-S. Su, M. Yajima, R2jags: Using R to Run 'JAGS', https://CRAN.R-project.org/package=R2jags, 2020. URL: https://CRAN.R-project.org/package=R2jags, accessed 2020-07-09.

H. Rue, L. Held, Gaussian Markov random fields: theory and applications, Chapman and Hall/CRC, 2005.

T. Janevic, E. Weber, F. M. Howell, M. Steelman, M. Krishnamoorthi, A. Fox, Analysis of state medicaid expansion and access to timely prenatal care among women who were immigrant vs us born, JAMA Network Open 5 (2022) e2239264–e2239264.

A. V. Bustamante, L. Félix-Beltrán, J. Nwadiuko, A. N. Ortega, Avoiding medicaid enrollment after the reversal of the changes in the public charge rule among latino and asian immigrants, Health Services Research 57 (2022) 195–203.

K. Korinek, K. R. Smith, Prenatal care among immigrant and racial-ethnic minority women in a

new immigrant destination: exploring the impact of immigrant legal status, Social science & medicine 72 (2011) 1695–1703.

L. Shi, L. A. Lebrun, J. Tsai, The influence of english proficiency on access to care, Ethnicity & health 14 (2009) 625–642.

Georgia Department of Community Health, Georgia Farmworker Health Program, 2025. URL: https://dch.georgia.gov/divisionsoffices/state-office-rural-health/farmworker-health-program, accessed: 2025-10-01.

J. S. Luque, G. Soulen, C. B. Davila, K. Cartmell, Access to health care for uninsured latina immigrants in south carolina, BMC health services research 18 (2018) 1–12.

S. Daudi, Choosing between healthcare and a green card: the cost of public charge, Emory LJ 70 (2020) 201.

H. Bernstein, D. Gonzalez, M. Karpman, S. Zuckerman, Amid confusion over the public charge rule, immigrant families continued avoiding public benefits in 2019, Washington, DC: Urban Institute 1 (2020) 1–20.

White House, Executive order on restoring faith in our legal immigration systems and strengthening integration and inclusion efforts for all americans. february 2, 2021.

S. S. Wang, S. Glied, C. Babcock, A. Chaudry, Changes in the public charge rule and health of mothers and infants enrolled in new york state's medicaid program, 2014–2019, American Journal of Public Health 112 (2022) 1747–1756.

Appendix A. Notation Table

Parameter Notation	Description
N_i	Population of immigrant mothers who gave birth in error-prone county i
M_j	Population of immigrant mothers who gave birth in gold standard county j
y_i	GDM counts for error-prone county i
z_j	GDM counts for gold standard county j
p_i	Error-prone probability of GDM for county i
θ_i	True probability for GDM in county i
μ	Global intercept
ρ	Global mixing parameter for spatially structured (u_i) and unstructured (v_i) random effects
ϕ	Global parameter for proportion of total variance attributed to spatially structured effects
X_i	Population at risk for county i
β	Covariate coefficients
v_i	Spatially unstructured random effect term for county i
σ_v^2	Variance of spatially unstructured random effect for county i
u_i	Spatially structured random effect (ICAR prior) for county i
τ	Smoothing parameter of the ICAR prior
λ_i	Sensitivity in county i
ω_i	Latent linear sensitivity predictor in county i
β_ω	Fixed effect for the average difference in the latent linear predictor ω_i for counties designated as gold standard (GS) compared to non-GS counties
v_ω	Spatially unstructured random effect term in ω_i across counties
σ_v^2	Variance of spatially unstructured random effect v_ω
u_i	Spatially structured random effect (ICAR prior) in ω_i across counties
D	Diagonal matrix containing the number of neighbors of each area on the diagonal
W	Adjacency matrix containing 1 for neighboring and 0 for non-neighboring counties

Appendix B. Full Model Specification and Conditional Distributions

Appendix B.1. Data Model

Appendix B.1.1. Observed True Counts

The observed true GDM counts are modeled using a binomial likelihood:

$$z_j \sim \text{Binomial}(M_j, \theta_i) \tag{B.1}$$

where θ_i represents the true prevalence of GDM.

Appendix B.1.2. Observed Error-Prone Counts

The observed error-prone GDM counts are modeled using a binomial likelihood:

$$y_i \sim \text{Binomial}(N_i, p_i) \tag{B.2}$$

where p_i represents the error-prone prevalence of GDM and θ_i is the true prevalence, modeled as:

$$p_i = \lambda_i \cdot \theta_i, \quad 0 < \lambda_i \leq 1 \tag{B.3}$$

Appendix B.2. Process Model

Appendix B.2.1. True GDM Risk

$$\text{logit}(\theta_i) = \mu + X_i'\beta + \rho\left(\sqrt{\frac{\phi}{1-\phi}}\right)u_i + (1-\rho)v_i, \quad i = 1,\ldots,159 \tag{B.4}$$

$$\mu \sim N(0,1) \tag{B.5}$$

$$\beta \sim N(0,\sigma_\beta^2) \tag{B.6}$$

$$v_i \sim N(0,\sigma_v^2) \tag{B.7}$$

$$u_i \sim N\left(0,[\tau_u(D-W)]^{-1}\right) \tag{B.8}$$

$$\rho \sim Beta(1,1) \tag{B.9}$$

$$\phi \sim Beta(1,1) \tag{B.10}$$

$$\tag{B.11}$$

Appendix B.2.2. Measurement Error Model

$$\lambda_i = \frac{1}{1+\exp(-\omega_i)} \tag{B.12}$$

$$\omega_i \sim N(\mu_\omega + \beta_{gs}\cdot \text{Indicator } (i \text{ in GS}) + u_\omega + v_\omega, 1) \tag{B.13}$$

$$\mu_\omega \sim N(0,0.01) \tag{B.14}$$

$$\beta_{gs} \sim \text{Truncated } N(1,1;[0,5]) \tag{B.15}$$

$$v_\omega \sim N(0,\sigma_{v_\omega}^2) \tag{B.16}$$

$$u_\omega \sim N\left(0,[\tau_\omega(D-W)]^{-1}\right) \tag{B.17}$$

$$\tag{B.18}$$

Appendix B.3. Full Conditional Distributions

Appendix B.3.1. Posterior of θ_i (True GDM Risk)

$$P(\theta_i \mid \text{data}) \propto P(y_i, z_j \mid N_i, \theta_i) P(\theta_i \mid \mu, X_i, u_i, \nu_i, \rho, \phi, \lambda_i) \tag{B.19}$$

Appendix B.3.2. Posterior of λ_i (Sensitivity Parameter)

$$P(\lambda_i \mid \text{data}) \propto P(y_i \mid p_i) P(\lambda_i \mid \omega_i) \tag{B.20}$$

$$\lambda_i \mid \omega_i \sim \text{Beta}(a_\lambda, b_\lambda) \tag{B.21}$$

Appendix B.3.3. Posterior of p_i (Error-Prone Risk)

$$P(p_i \mid \text{data}) \propto P(y_i \mid p_i) P(p_i \mid \lambda_i, \theta_i) \tag{B.22}$$

Appendix B.3.4. Posterior of Spatial Effects (u_i and ω_i)

$$P(u_i \mid \text{neighbors}) \propto \text{MVN}(\bar{u}_i, \sigma_u^2) \tag{B.23}$$

$$P(\omega_i \mid \text{neighbors}) \propto \text{MVN}(\bar{\omega}_i, \tau_\omega^{-1}) \tag{B.24}$$

Appendix B.3.5. Posterior of Regression Coefficients β

$$\beta \mid \text{data} \sim \text{MVN}(\hat{\beta}, V_\beta) \tag{B.25}$$