Modeling Complex Survey Data: a Case Study of International Health Surveillance Surveys

Timothy Raxworthy[1], Yajuan Si[2] and Grace Chung[3]

[1]University of Michigan, Ann Arbor, MI 48104 USA, raxworth@umich.edu
[2]Institute for Social Research, 426 Thompson St, Ann Arbor, MI 48104 USA, yajuan@umich.edu
[3]University of New Mexico, 2300 Tucker Ave NE, Albuquerque, NM 87131 USA, gchung@salud.unm.edu

Abstract

Model-based approaches to inference are common in the presence of complex survey data. Although statistical modeling is an often necessary approach for analyzing data, there is no firm consensus as to how these analyses should handle sampling weights. Using a case study of international health surveillance surveys, this paper examines the roles of weights in the generalized linear models (GLMs) and generalized linear mixed-effect models (GLMMs). We considered two different ways of including weights with model estimates: using weighted likelihood functions for model fitting and weighted average values of individual predictions. We compared GLM and GLMM estimates as well as unweighted and weighted variants of these models. We found that including weights in the model fitting processes does not substantially change the model parameter estimates and predictions. The difference between weighted and unweighted descriptive statistics is more pronounced than that of the model parameter estimates. We recommend comparing the weighted and unweighted descriptive summaries as a standard analysis routine in practice.

Keywords: multilevel modeling, complex sample survey, weights, diabetes prevalence.

1 Introduction

Most researchers in the social sciences and public health use sample survey data for finite population inference and account for complex sample design features if they are related to the survey outcome (Si, Lee, and Heeringa, 2024). When the sampling design features are informative, e.g., the selection leads to the sample distribution of the quantity of interest deviating from the underlying population distribution, appropriate analysis methods are necessary to adjust for the sample discrepancy. Despite this general recommendation, there have been historical debates on whether survey weights are necessary when fitting statistical models. Survey weights often require special treatments to meet researchers' analytic goals and are sometimes considered to be a nuisance when they only inflate the variance of an estimate without changing the point estimate itself. In addition to model specification, the various ways handling weights can also be due to factors such as researchers' familiarity with using survey weights, the quality of documentation on how weights are constructed, and the availability of auxiliary information about the population. Using a case study of international health surveillance surveys, this paper examines the roles of weights in the generalized linear models (GLMs) and generalized linear mixed-effect models (GLMMs).

Motivated by the need of meeting the World Health Organization's (WHO) recommended diabetes targets (Gregg et al., 2023), we aim to estimate and compare diabetes prevalence between countries. We use the WHO Stepwise Approach to Surveillance (STEPS) surveys (Riley et al., 2016), which are cross-sectional probability sample surveys conducted in more than 100 countries and collect various health risk indicators to provide population level estimates. Most STEPS surveys have implemented

multistage stratified data collections. First, enumeration areas are selected as primary sampling units (PSUs) and then households are selected as secondary sampling units. The final stage randomly selects eligible individuals from each household. Following the design of the STEPS survey with multistage stratified sampling, every individual i in the sample was assigned a survey weight w_i. To properly estimate the sampling variance, it is necessary to account for stratification, PSU clustering, and weights in the analysis. Further, we are also interested in assessing whether the sample survey design features affect the estimation of diabetes prevalence across multiple countries.

The paper structure is organized as follows. Section 1.1 describes the data and measures used in this study in detail. Section 2 introduces the model-based inference approach. The results from different methods in Section 3 are then compared. Section 4 summarizes the main takeaways from the study.

1.1 Data and Measures

We used a subset of the WHO's STEPS data in this analysis, which were collected from 10 different countries between 2015 and 2016. This dataset was processed, resulting in 26,752 individuals. We define a case with diabetes as having any of the following: (1) a fasting plasma glucose of 7.0 mmol/L or higher, (2) hemoglobin A1c (HbA1c) level of 6.5% or higher, or (3) self-reported use of glucose-lowering medication or use of insulin or oral hypoglycemic drugs. This definition of diabetes is used in the Global Monitoring Framework for Non Communicable Diseases (Gregg et al., 2023). Once these cases are computed, a binary indicator variable for diabetes is then defined and used as the dependent variable in the analysis. Eligible participants who have been assigned survey weights are included. The person level covariates used in predicting our outcome are body mass index (BMI), age, sex, and highest completed education level.

Table 1: Description of Each Country's Sample and Sampling Weights

Country	Sample Size	Sum of Weights	Mean	SE
Algeria	6,393	25,888,236	4,049.47	1,316.80
Benin	5,073	2,441,103	481.20	1,686.59
Brunei	2,018	102,824	50.95	43.13
Ethiopia	9,800	34,097,395	3,479.33	3,707.59
Guyana	1,178	203,440	172.70	144.86
Iraq	4,071	14,942,707	3,670.53	3,569.90
Kiribati	2,156	57,561	26.70	59.80
Nauru	1,387	4,212	3.04	0.33
Solomon Is.	2,522	341,164	135.28	123.54
Vietnam	3,758	78,831,165	20,976.89	17,094.49

Table 1 includes further detailed information regarding each country sample. The 'Sample Size' column refers to the total number of eligible participants assigned a sampling weight and the 'Sum of Weights' is the calculated total sum of the sampling weights. The 'Mean' column is the mean value of the sampling weights and 'SE' column is the standard error of the weights.

There were missing values for some respondents' BMI and education. The amount of item nonresponse is less than 5% of the records. We assumed missing at random and used multiple imputation to fill in missing item values with a proportional odds model accounting for the order of the BMI and education categories. Along with these variables, other covariates used in multiple imputation were the indicator variable for a diabetic case, age, and sex. We used one imputed dataset for simplicity,

even though multiple completed datasets could have been pooled for analysis via combining rules.

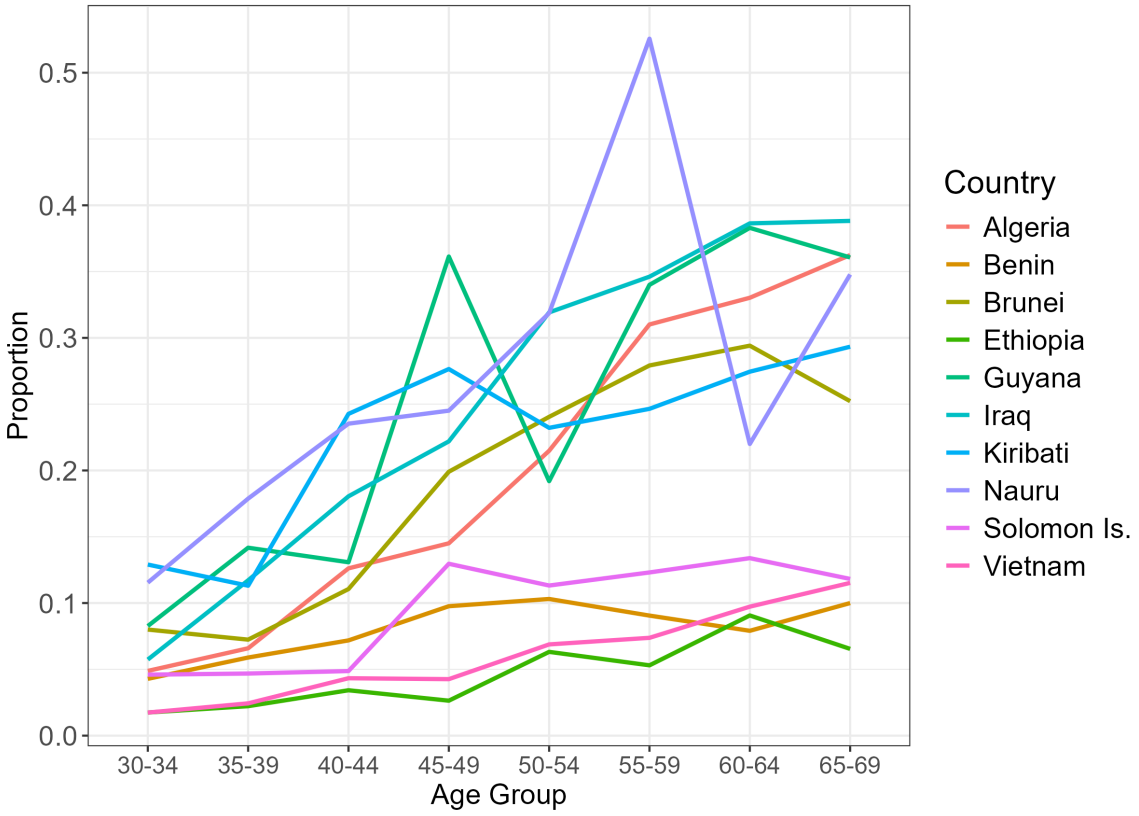

Figure 1: Observed proportions of individuals with diabetes by age group and country.

Figure 1 displays how the observed proportions of individuals with diabetes change across age groups for each country included in our analysis. The generally increasing trends are similar between countries but the changing rates over time are accentuated for Algeria, Brunei, Guyana, Iraq, Kiribati and Nauru. Table 1 shows that the sample sizes across countries also varies largely. Further, the likely cause for the spikes shown at specific age cohorts is due to small sample size for those ages.

2 Modeling Approach

We use the individual level data to model the probability of having diabetes in a logistic regression. First, we define the outcome variable as:

$$y_{ij} = \begin{cases} 1 & \text{if individual } i \text{ in country } j \text{ has the disease,} \\ 0 & \text{if individual } i \text{ in country } j \text{ does not have the disease,} \end{cases}$$

for $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$, where n_j is the total number of individuals, and J is the total number of countries. Our GLM model is specified as

$$log \left\{ \frac{Pr(y_{ij} = 1 | \boldsymbol{X'_{ij}})}{Pr(y_{ij} = 0 | \boldsymbol{X'_{ij}})} \right\} = \boldsymbol{X'_{ij}} \boldsymbol{\beta}, \tag{1}$$

where $\boldsymbol{X'_{ij}}$ denotes the person-level covariates, including BMI, age, sex, and education, and country indicators.

In the GLMM, we include country-level and PSU-level random intercepts to borrow information across countries and clusters, stabilize estimates for small countries and account for the PSU clustering effects with the following specification:

$$log\left\{\frac{Pr(y_{ij}=1|\boldsymbol{X'},u_j,v_{k[i]})}{Pr(y_{ij}=0|\boldsymbol{X'},u_j,v_{k[i]})}\right\}=\boldsymbol{X'\beta}+u_j+v_{k[i]}, \tag{2}$$

where $\boldsymbol{X'}$ denotes the person-level covariates, u_j's are the country-varying effects, and $v_{k[i]}$'s are the PSU-varying effects, where $k[i]$ is the PSU index k that individual i is assigned to, $u_j \overset{iid}{\sim} N(0,\sigma_u^2)$, $v_k \overset{iid}{\sim} N(0,\sigma_v^2)$, with both random effects assumed to be independent, identically and normally distributed with a mean of 0 and variance σ_u^2 and σ_v^2, respectively. The intra-country correlation (ICC) is also measured:

$$\rho=\frac{\sigma_u^2}{\sigma_u^2+\sigma_v^2+\frac{\pi^2}{3}}.$$

Although we considered different interactions between the categorical variables and country level predictors, we did not include additional covariates because of model estimation problems. We consider two different ways to account for weights in estimating the country-specific diabetes prevalence as the proportion of people with the characteristics described in 1.1.

1) Using a weighted average of the estimated predicted probability $\hat{p}_{ij}=Pr(y_{ij}=1|\boldsymbol{X'_{ij}})$ of having diabetes for individual i in country j based on Model (1) conditional on $\boldsymbol{X'_{ij}}$. For Model (2) this is expressed as $Pr(y_{ij}=1|\boldsymbol{X'},u_j,v_{k[i]})$. The weighted average of the predicted probabilities is given based on the Hájek estimator (Hájek, 1971). For country j, the weighted prevalence is given by

$$\hat{\theta}_j=\frac{\sum_{i\in s_j} w_i * \hat{p}_{ij}}{\sum_{i\in s_j} w_i}, \tag{3}$$

where s_j is the sample of individuals in country j.

2) Including weights in the model fitting processes of either the GLM in (1) or the GLMM in (2). We use the pseudo maximum likelihood (PML) estimation to obtain the parameter estimates that maximum the weighted likelihood function, where each individual's likelihood is powered by the corresponding weight value (Skinner, 1989). The weighted GLM likelihood $l_{WGLM}(\cdot)$ and weighted GLMM likelihood $l_{WGLMM}(\cdot)$ are as below.

$$l_{WGLM}(y,\boldsymbol{X'},\beta,w)=\prod_{i=1}^{n}\left[\left(\frac{\exp(\boldsymbol{X'}_{ij}\beta)}{1+\exp(\boldsymbol{X'}_{ij}\beta)}\right)^{y_{ij}}\left(\frac{1}{1+\exp(\boldsymbol{X'}_{ij}\beta)}\right)^{(1-y_{ij})}\right]^{w_i} \tag{4}$$

$$l_{WGLMM}(y,\boldsymbol{X'},\beta,u,v,w)=\prod_{i=1}^{n}\left[\left(\frac{\exp(\boldsymbol{X'\beta}+u_j+v_{k[i]})}{1+\exp(\boldsymbol{X'\beta}+u_j+v_{k[i]})}\right)^{y_{ij}}\left(\frac{1}{1+\exp(\boldsymbol{X'\beta}+u_j+v_{k[i]})}\right)^{(1-y_{ij})}\right]^{w_i}. \tag{5}$$

We normalize the weights when fitting the GLMM. Rabe-Hesketh and Skrondal, 2006 show that the weighted likelihood function requires weights at each level of the data hierarchy. Based on the STEPS survey design, different countries independently conducted the surveys, and there was no random selection of countries. To effectively pool estimates across countries, we scaled the person-level weights using method 2 described in Pfeffermann et al., 1998. The weight adjustment scales the weights w_i of individuals in country j, for $i \in s_j$, by adjusting the sum to be equal to the sample size of each country n_j. The adjustment factor a_j for individuals in country j can be expressed as: $a_j=\frac{n_j}{\sum_{i\in s_j} w_i}$, the product of which and the weight w_i will be used in the pseudo maximum likelihood (PML) estimation based on the GLMM. This scaling method has been described as performing better

in simulations where the design is considered informative (Pfeffermann et al., 1998).

3 Model Inference

We fit GLM and GLMM models, both unweighted and weighted, to predict the diabetes prevalence with age (30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, and 65-69), BMI ($<$25, $\leq$25 & $<$30, and $\geq$30), education (no education, primary school, high school [HS] & above), sex (male, and female), and country indicators (Algeria, Benin, Brunei, Ethiopia, Guyana, Iraq, Kiribati, Nauru, Solomon Islands, and Vietnam).

We account for stratification and clustering in the standard error estimation for both models. We use analytic variance estimation via Taylor series linearization by default in the R *survey* package (Lumley, 2024) and Stata (StataCorp, 2025), i.e., defining the complex survey design object including strata codes, PSU codes, and sampling weights (for weighted estimates). To obtain country-specific estimates, we apply the unconditional approach for variance estimation and takes the full complex sample design into account when analyzing subpopulations (Heeringa, West, and Berglund, 2017). When summarizing the model predictions, the complex survey design features (PSUs, strata and weights) are accounted for to obtain design-based estimates using expression (3).

3.1 Model Estimation

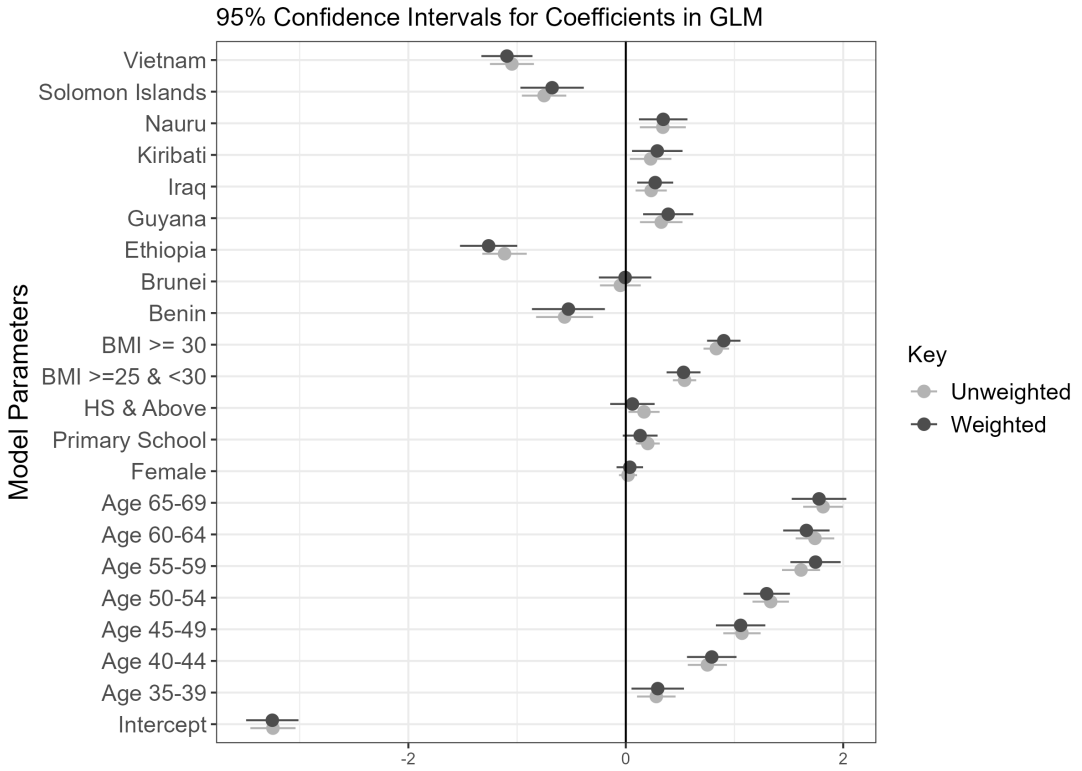

Figure 2: Comparison of coefficient estimates and 95% confidence intervals for the true coefficients between unweighted GLM 1 and weighted GLM 4.

First, we fit the unweighted GLM in (1) and weighted GLM in (4). The reference categories for each of the predictors are those persons who are aged 30-34, male, have received no formal education, have a BMI $<$ 25 and reside in Algeria which is the first country in alphabetical order. The model coeffi-

cient estimates together with the corersponding 95% confidence intervals for the true coefficients are presented in Figure 2 . Including weights in model fitting tends to increase the variance of coefficient estimation, resulting in wider confidence intervals as shown in Figure 2.

When comparing the coefficient point estimates of the weighted model versus the unweighted model, the weights slightly change our interpretation of the coefficients associated with each country. For example, across all countries, the weighted estimate's 95% confidence interval for the effect of primary school contain zero when its corresponding unweighted estimate does not. For all countries, the age coefficient estimates in the weighted model are either lower or higher than those in the unweighted model, changing without any apparent pattern. The opposite trend is the case for coefficients associated with education where the weights are reducing the magnitude for the effects of both age categories. These subtle changes are due to the differences in the maximum likelihood estimates when the weights are added in the GLM model fitting. The role of weights depends on the model specification and its dependency on the sample design features. Therefore, fitting both unweighted and weighted models is helpful when analyzing complex survey samples to gain more insight into how the weights interact with the model fitting process.

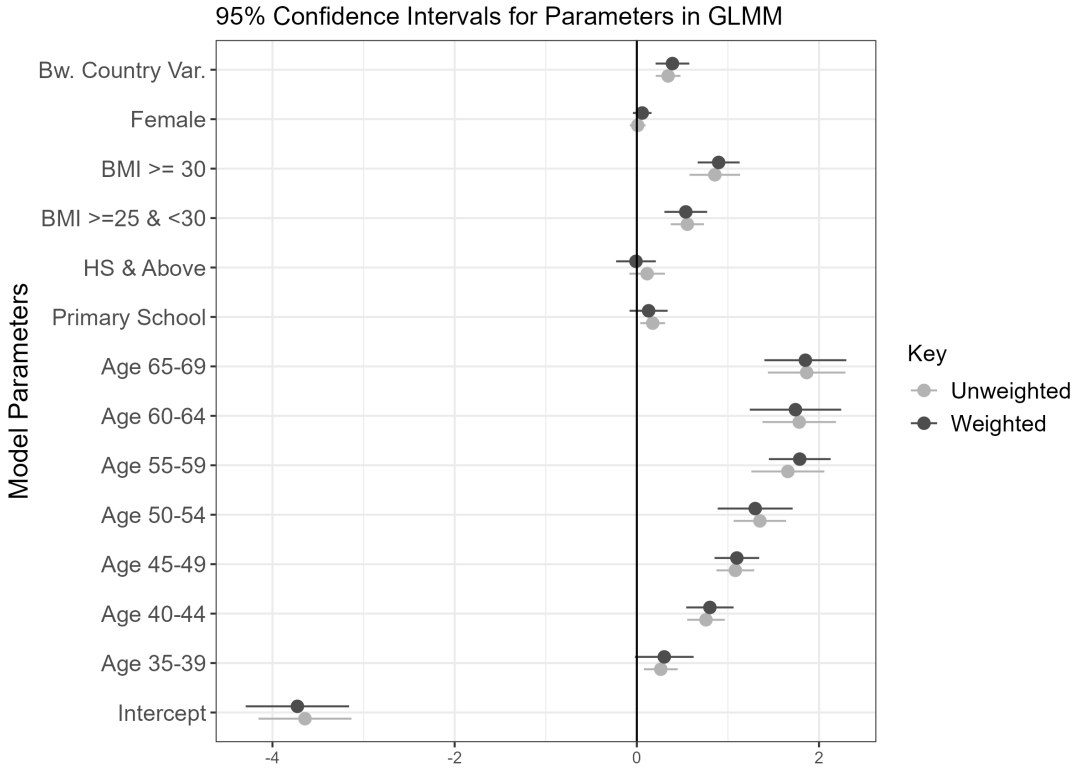

Figure 3: Comparison of coefficient estimates and 95% confidence intervals between unweighted and weighted generalized linear mixed-effect models.

Next we fit the unweighted GLMM in (2) and weighted GLMM in (5). The ICC measuring the intra-country similarity is estimated to be 0.084 for the unweighted model and 0.094 in the weighted model. Figure 3 compares the estimated model coefficients for the weighted and unweighted GLMM models. In general, the model coefficients are similar between both models, although the weighted GLMM model estimates have larger variances.

In sum, with either GLM or GLMM, including the weights in model fitting does not substantially change the model parameter estimates.

Table 2: Model predicted (Pred), weighted model predicted (Weighted pred), observed (Obs) and weighted observed (Weighted Obs.) diabetes prevalence in 10 countries with standard error values reported in parentheses.

Country	Obs	Pred		Pred	
		GLM	Weighted GLM	GLMM	Weighted GLMM
Algeria	.166 (.007)	.166 (.002)	.163 (.002)	.158 (.004)	.154 (.004)
Benin	.073 (.008)	.073 (.001)	.075 (.001)	.066 (.003)	.067 (.003)
Brunei	.184 (.011)	.184 (.003)	.182 (.003)	.175 (.006)	.168 (.008)
Ethiopia	.036 (.003)	.036 (.001)	.032 (.001)	.032 (.001)	.027 (.001)
Guyana	.234 (.014)	.234 (.004)	.239 (.004)	.219 (.005)	.223 (.005)
Iraq	.215 (.009)	.215 (.002)	.220 (.002)	.204 (.004)	.205 (.004)
Kiribati	.208 (.013)	.208 (.003)	.213 (.003)	.201 (.009)	.201 (.008)
Nauru	.233 (.015)	.233 (.004)	.232 (.004)	.228 (.013)	.227 (.013)
Solo. Is.	.082 (.007)	.082 (.001)	.086 (.001)	.076 (.004)	.078 (.006)
Vietnam	.053 (.004)	.053 (.001)	.050 (.001)	.047 (.001)	.044 (.001)

Country	Weighted Obs.	Weighted pred		Weighted pred	
		GLM	Weighted GLM	GLMM	Weighted GLMM
Algeria	.153 (.007)	.156 (.002)	.153 (.002)	.147 (.004)	.144 (.004)
Benin	.069 (.011)	.067 (.003)	.069 (.003)	.062 (.005)	.063 (.007)
Brunei	.164 (.014)	.164 (.003)	.164 (.003)	.159 (.006)	.154 (.008)
Ethiopia	.030 (.003)	.035 (.001)	.030 (.001)	.029 (.001)	.026 (.001)
Guyana	.220 (.017)	.215 (.005)	.220 (.005)	.201 (.005)	.204 (.007)
Iraq	.217 (.011)	.212 (.003)	.217 (.003)	.204 (.005)	.205 (.005)
Kiribati	.211 (.014)	.207 (.007)	.211 (.007)	.199 (.007)	.206 (.011)
Nauru	.235 (.015)	.236 (.005)	.235 (.005)	.231 (.013)	.230 (.013)
Solo. Is.	.079 (.010)	.076 (.001)	.079 (.001)	.071 (.004)	.073 (.006)
Vietnam	.046 (.004)	.049 (.001)	.046 (.001)	.043 (.001)	.040 (.001)

3.2 Model Prediction

Using both the weighted and unweighted variants of the GLM and GLMM described above, we predict the response probability for each individual and estimate prevalence by country. The top table in Table 2 compares the observed prevalence as the simple proportion of diabetic cases divided by the total sample size with the predictions from the GLM in (1), weighted GLM in (4), GLMM in (2), and weighted GLMM in (5). The bottom table in Table 2 applies weights to the individual probabilities based on (3) and presents the weighted observation using expression (3) and prediction values. Based on Table 2, model-based predictions have lower standard errors than the observed prevalence of each country. We have omitted the model-based error in predicting the probabilities, but the prediction variability is smaller than the sampling error. The country-level prevalence values calculated as the average of individual predictions from the GLM are the same as the observed summaries, which is as expected because the GLM model includes the fixed effects of countries. Nevertheless, the GLMM includes the random effects of countries, and partial pooling across countries yields predicted summaries different from the observed values. Comparing GLM and GLMM before and after weighting, we see that using weighted likelihood estimation does not substantially change the predicted values or standard errors. The predictions based on GLMM have more variability than those based on GLM, probably due to the inclusion of random effects.

Applying weights to the individual probabilities based on (3) does change the point estimates across countries and increases the standard errors. This is true for both observed and predicted values from all models. The weighted predictive averages based on the weighted GLM are the same as the weighted observed summaries but the standard errors are lower. The reason for the weighted predictive averages being the same is because the weights have been normalized to equal the sum of the sample size within each country. Overall, this shows that using a weighted average of the predicted probabilities generates larger influences than including weights in the model fitting processes. Weights are more influential for descriptive summaries than model estimates, which is consistent with the literature findings, e.g., Si, Lee, and Heeringa, 2024.

4 Conclusion

In this study, we used a case study with international health surveys to assess the role of survey weights in model inference and prevalence estimation. We considered two different ways of including weights with model estimates: using weighted likelihood functions for model fitting and weighted average values of individual predictions. We compared GLM and GLMM estimates as well as unweighted and weighted variants of these models. We found that including weights in the model fitting processes does not substantially change the estimated model coefficients and predictions. The difference between weighted and unweighted prevalence summaries is more pronounced than that of the model parameter estimates. We recommend comparing the weighted and unweighted descriptive summaries as a standard analysis routine in practice.

Finally, our empirical comparisons cannot be validated without knowing the gold-standard or true values. When comparing these diabetes estimates to those published in other sources, there may be some cases where the unweighted estimates are closer to estimates found by experts in these fields, while other cases have weighted estimates with closer comparisons. This dilemma demonstrates the need to work with experts in the topic of analysis who can properly evaluate survey estimates beyond the statistical component. As pointed out above, we only use one set of predicted probabilities for the point estimates and omitted the prediction error due to model fitting. The model-based error is smaller than the sampling error. Future work would be needed to develop practical methods that account for both modeling and sampling error, such as using Monte Carlo simulation.

5 Data and Software

This paper uses data from the Algeria 2016 (Ministry of Health (Algeria), Population and Hospital Reform, and World Health Organization (WHO), 2017), Benin 2015 (Ministry of Health (Benin), and World Health Organization (WHO), 2015), Brunei Darussalam 2015-2016 (Ministry of Health (Brunei), and World Health Organization (WHO), 2016), Ethiopia 2015 Ethiopia Public Health Institute, Federal Ministry of Health (Ethiopia) and World Health Organization (WHO), 2016, Guyana

2016 (Pan American Health Organization (PAHO), Ministry of Public Health of Guyana, and the Bureau of Statistics (Guyana), 2019), Iraq 2015 (Ministry of Health (Iraq), Ministry of Planning (Iraq), World Health Organization (WHO), 2015), Kiribati 2015-2016 (Ministry of Health and Medical Services (Kiribati), World Health Organization (WHO), 2015), Nauru 2015 (Ministry of Health (Nauru), World Health Organization (WHO), 2016), Solomon Islands 2015 (Ministry of Health (Solomon Islands), World Health Organization (WHO), 2020), and Viet Nam 2015 (Ministry of Health (Vietnam), World Health Organization (WHO), 2016) STEPS surveys. These surveys were implemented by the agen-

cies listed in each citation along with support by the World Health Organization. These datasets are available upon request from the https://extranet.who.int/ncdsmicrodata/index.php/homeWHO NCD Microdata Repository.

The svyglm function in the https://cran.r-project.org/package=surveysurvey package for R software was used to fit GLMs (Lumley, 2010). The https://www.stata.com/manuals/memelogit.pdfmelogit function within Stata software was used to fit the GLMMs.
The https://cran.r-project.org/package=micemice package of R software was used for multiple imputation (Van Buuren and Groothuis-Oudshoorn, 2011).

References

Ethiopia Public Health Institute, Federal Ministry of Health (Ethiopia) and World Health Organization (WHO) (2016). *Ethiopia Steps Report On Risk Factors for Chronic Noncommunicable Diseases and Prevalence of selected NCDs*. World Health Organization STEPS Survey Repository.

Gregg, E. W., J. Buckley, M. K. Ali, J. Davies, D. Flood, and et al (2023). "Improving health outcomes of people with diabetes: target setting for the WHO Global Diabetes Compact". In: *Lancet* 401.10384, pp. 1302–1312.

Hájek, J. (1971). "Comment on "An Essay on the Logical Foundations of Survey Sampling, Part One"". In: *The Foundations of Survey Sampling*.

Heeringa, S. G., B. T. West, and P. A. Berglund (2017). *Applied Survey Data Analysis*. 2nd. CRC Press.

Lumley, T. (2010). *Complex Surveys: A Guide to Analysis Using R*. John Wiley and Sons.

— (2024). *survey: analysis of complex survey samples*. R package version 4.4.

Ministry of Health (Algeria), Population and Hospital Reform, and World Health Organization (WHO) (2017). *Algeria STEPS Noncommunicable Disease Risk Factors Survey 2016–2017*. World Health Organization STEPS Survey Repository.

Ministry of Health and Medical Services (Kiribati), World Health Organization (WHO) (2015). *Kiribati NCD Risk Factors STEPS REPORT 2015-2016*. World Health Organization STEPS Survey Repository.

Ministry of Health (Benin), and World Health Organization (WHO) (2015). *Benin STEPS Noncommunicable Disease Risk Factors Survey 2015*. World Health Organization STEPS Survey Repository.

Ministry of Health (Brunei), and World Health Organization (WHO) (2016). *Brunei STEPS Noncommunicable Disease Risk Factors Survey 2015-2016*. World Health Organization STEPS Survey Repository.

Ministry of Health (Iraq), Ministry of Planning (Iraq), World Health Organization (WHO) (2015). *Noncommunicable Diseases Risk Factors STEPS Survey Iraq 2015 Report*. World Health Organization STEPS Survey Repository.

Ministry of Health (Nauru), World Health Organization (WHO) (2016). *Nauru NCD RIsk Factors STEPS Report 2015-2016*. World Health Organization STEPS Survey Repository.

Ministry of Health (Solomon Islands), World Health Organization (WHO) (2020). *Solomon Islands NCD Risk Factors STEPS REPORT 2015*. World Health Organization STEPS Survey Repository.

Ministry of Health (Vietnam), World Health Organization (WHO) (2016). *National Survey on the Risk Factors of Non-Communicable Diseases (STEPS) Viet Nam 2015*. World Health Organization STEPS Survey Repository.

Pan American Health Organization (PAHO), Ministry of Public Health of Guyana, and the Bureau of Statistics (Guyana) (2019). *Pan American STEPS Risk Factor Surveillance Survey Guyana 2016 Report*. World Health Organization STEPS Survey Repository.

Pfeffermann, D., C. J. Skinner, D. J. Holmes, H. Goldstein, and J. Rasbash (1998). "Weighting for unequal selection probabilities in multilevel models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60.1, pp. 23–40.

Rabe-Hesketh, S. and A. Skrondal (2006). "Multilevel modelling of complex survey data". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169.4, pp. 805–827.

Riley, L., R. Guthold, M. Cowan, S. Savin, L. Bhatti, T. Armstrong, and R. Bonita (2016). "The World Health Organization STEPwise Approach to Noncommunicable Disease Risk-Factor Surveillance: Methods, Challenges, and Opportunities". In: *American Journal of Public Health* 106.1, pp. 74–78.

Si, Y., S. Lee, and S. G. Heeringa (2024). "Population weighting in statistical analysis". In: *JAMA Internal Medicine* 184.1, pp. 98–99.

Skinner, C. J. (1989). "Domain means, regression and multi-variate analysis". In: *Analysis of Complex Surveys*. Ed. by C. J. Skinner, D. Holt, and T. M. F. Smith. Wiley, pp. 59–88.

StataCorp (2025). *Stata Statistical Software: Release 19*.

Van Buuren, S. and K. Groothuis-Oudshoorn (2011). "mice: Multivariate Imputation by Chained Equations in R". In: *Journal of Statistical Software* 45.3, pp. 1–67.