iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

ESTIMATING DIABETES PREVALENCE TRENDS USING WEIGHTED NATIONAL SURVEY DATA AND CONFIDENCE INTERVAL COMPUTATIONS

Tahiru Mahama
Department of Mathematical Sciences, The University of Texas at El Paso, USA

ABSTRACT
Diabetes mellitus has emerged as a critical public health concern globally, with prevalence rates escalating due to lifestyle changes, aging populations, and urbanization. To effectively address and mitigate the burden of diabetes, it is essential to understand its temporal prevalence trends across diverse populations. This study investigates diabetes prevalence trends by analyzing weighted data from national health surveys, which provide representative insights into population-level health indicators. Employing complex survey design adjustments, we ensure that estimates reflect the true distribution of diabetes across age, sex, ethnicity, and geographic subpopulations. The core analytical approach involves the application of weighted prevalence estimation, which accounts for unequal probabilities of selection, non-response, and post-stratification. By incorporating confidence interval computations, specifically through Taylor series linearization and bootstrap resampling, we quantify the statistical uncertainty surrounding prevalence estimates. These intervals are crucial for interpreting trends over time and determining whether observed differences are statistically significant or within the margin of error. Our findings reveal nuanced patterns in diabetes prevalence, including rising rates in younger adult groups and persistently high levels in elderly populations. Stratified trend analyses indicate demographic shifts in disease burden, suggesting that interventions must be tailored to emerging at-risk groups. Additionally, visualizing prevalence with corresponding confidence bounds enhances clarity in temporal trend assessments and aids in public health planning. This study highlights the value of integrating robust statistical techniques with nationally representative health survey data to generate precise, generalizable insights. The use of weighted data and reliable confidence intervals ensures that health surveillance efforts are both accurate and informative for policymakers and healthcare providers aiming to reduce diabetes incidence and its complications.

Keywords:
Diabetes prevalence; Weighted survey data; National health surveys; Confidence intervals; Trend analysis; Public health surveillance

1. INTRODUCTION

1.1 Global Burden of Diabetes and Public Health Impact

Diabetes mellitus has emerged as one of the most significant non-communicable diseases (NCDs) affecting global health, with rising prevalence across both developed and developing nations. In 2021, it was estimated that over 537 million adults worldwide were living with diabetes, and this number is projected to rise to 643 million by 2030 [1]. This upward trajectory reflects changes in diet, urbanization, sedentary lifestyles, and increased life expectancy—all of which contribute to elevated metabolic risk at the population level [2].

Beyond its direct health impact, diabetes imposes a substantial economic burden on healthcare systems. Direct costs include expenditures on insulin, glucose monitoring, hospitalizations, and complications such as diabetic nephropathy and retinopathy, while indirect costs arise from lost productivity, disability, and premature mortality [3]. In low- and middle-income countries, where healthcare systems often struggle with dual burdens of infectious and chronic diseases, diabetes care remains highly under-resourced [4].

Additionally, diabetes significantly increases the risk of cardiovascular disease, stroke, and lower limb amputation, thereby contributing to years of life lost and disability-adjusted life years (DALYs) [5]. The growing prevalence also places strain on health workers, particularly in primary care settings where early detection and routine management are essential but often lacking.

The global burden of diabetes underscores the urgency of integrating chronic disease prevention and management into public health agendas. Multisectoral efforts—including public awareness, food policy regulation, and community-based screening—are essential to reduce incidence and improve outcomes [6]. However, effective

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

planning for these interventions requires accurate national-level data to assess trends, identify at-risk populations, and allocate resources equitably [7].

1.2 Importance of National Surveys in Chronic Disease Surveillance

Nationally representative health surveys play a foundational role in the surveillance and management of chronic diseases such as diabetes. These surveys provide comprehensive data on disease prevalence, behavioral risk factors, healthcare access, and treatment patterns across various demographic subgroups [8]. By employing stratified sampling and standardized data collection protocols, national surveys ensure that findings are generalizable and comparable over time and across regions.

One of the most widely recognized initiatives is the WHO STEPwise approach to NCD risk factor surveillance (STEPS), which provides standardized modules for assessing self-reported and biometric indicators of diabetes and related conditions [9]. Similarly, national demographic and health surveys (DHS) and behavioral risk factor surveillance systems (BRFSS) have been used extensively to monitor population health and inform policy design [10].

Such surveys are especially critical in low-resource settings, where clinical registries are often incomplete or non-existent. They help fill data gaps by capturing information from underserved or rural populations that may otherwise be excluded from facility-based studies [11]. Moreover, national surveys support longitudinal tracking of NCD targets as defined in global action plans, such as the Sustainable Development Goals (SDGs) and the WHO Global NCD Monitoring Framework [12].

Importantly, data from these surveys not only guide public health programming but also inform academic research, funding priorities, and global comparisons. As the burden of diabetes continues to rise, strengthening national survey systems and expanding their reach remains a vital priority for chronic disease control at scale [13].

1.3 Objectives: Accurate Prevalence Estimation, Weighting, and Uncertainty Quantification

Accurately estimating the national prevalence of diabetes is essential for effective health system planning and resource allocation. Given the complex sampling designs of most national health surveys—often involving stratification, clustering, and unequal probabilities of selection—appropriate analytical techniques are required to derive valid population-level estimates [14]. Without correction for these design effects, raw estimates may be misleading and not truly reflective of national or regional disease burdens.

A primary objective of survey analysis is to apply **sampling weights** to adjust for differential selection probabilities and non-response. These weights ensure that each observation contributes appropriately to the overall estimate based on its representation in the target population [15]. For example, if rural regions are oversampled in a survey, failing to apply weights would overstate their influence in prevalence calculations.

Another important aim is the estimation of **standard errors and confidence intervals** that account for the survey's complex design. Simple random sampling assumptions lead to underestimation of uncertainty, which may result in spurious inferences or policy misdirection [16]. Techniques such as Taylor series linearization or replication methods (e.g., jackknife, bootstrap) are typically used to calculate design-adjusted variance estimates [17].

Uncertainty quantification is not just a statistical exercise—it is integral to transparent policy communication. Presenting decision-makers with weighted prevalence rates and confidence intervals allows for informed prioritization and targeted intervention planning [18]. Thus, the analytical goals of weighting and error estimation are central to leveraging national survey data for chronic disease surveillance, particularly in estimating the true scope and impact of diabetes across diverse populations [19]. Given the sampling complexity and national relevance of chronic disease survey data, it is methodologically essential to implement **weighted survey analysis**. This ensures that diabetes prevalence estimates are not only statistically valid but also truly representative of national population structures. Section 2 will explore the core design elements of complex surveys and the analytical tools needed to produce accurate, weighted outcomes.

2. NATIONAL HEALTH SURVEYS AND SAMPLING DESIGN

2.1 Overview of Major National Surveys (NHANES, BRFSS, STEPS)

National health surveys have long served as the backbone of population-level chronic disease surveillance, especially for conditions such as diabetes. Among the most prominent and widely utilized tools are the National Health and Nutrition Examination Survey (NHANES), the Behavioral Risk Factor Surveillance System (BRFSS), and the WHO STEPwise approach to NCD surveillance (STEPS). Each of these programs provides crucial insights into the prevalence, trends, and determinants of diabetes and related risk factors across diverse populations [6].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

NHANES, conducted in the United States, is a unique and comprehensive survey that combines interviews with physical examinations and laboratory measurements. It uses a stratified multistage probability design to ensure representativeness of the non-institutionalized civilian population [7]. NHANES is particularly valuable because it offers objective biomarkers—including fasting glucose and HbA1c levels—that enable accurate classification of diabetes and prediabetes status across subgroups.

BRFSS, also implemented in the United States, is the largest continuously conducted health survey globally. It relies on telephone-based interviews to collect self-reported data on health behaviors, preventive services, and diagnosed conditions [8]. Although BRFSS does not include laboratory tests, it remains indispensable for monitoring diabetes prevalence, behavioral risk factors, and access to care over time. The survey's wide reach and annual frequency make it a cornerstone of state-level diabetes surveillance and public health evaluation.

The WHO's STEPS survey is an internationally standardized tool developed to assist countries in gathering consistent and comparable data on non-communicable disease risk factors, including diabetes [9]. STEPS surveys follow a sequential approach: Step 1 collects questionnaire-based data, Step 2 includes physical measurements, and Step 3 adds biochemical assessments. This tiered design enables flexible implementation depending on national resources.

Collectively, these surveys have laid the foundation for evidence-based chronic disease control. Their structured methodologies, recurring intervals, and scalable formats have supported decades of health planning, resource allocation, and progress evaluation in diabetes prevention and care [10].

2.2 Complex Survey Design: Stratification, Clustering, and Unequal Probabilities

Understanding the design of complex health surveys is crucial for proper interpretation and statistical analysis. Unlike simple random sampling, national health surveys employ intricate sampling strategies that account for the heterogeneity of target populations. Three key design elements—stratification, clustering, and unequal probability sampling—are commonly used to enhance efficiency and representativeness [11].

Stratification involves dividing the population into distinct subgroups (strata) based on characteristics like geographic region, urban-rural status, or socioeconomic level. Sampling is then conducted independently within each stratum. This ensures that all relevant subpopulations are adequately represented in the final dataset and allows for precise estimates within each stratum [12].

Clustering refers to the grouping of individuals into primary sampling units (PSUs), such as households, census blocks, or districts, before selection. Clustering is typically used to reduce fieldwork costs and logistical complexity, especially in geographically dispersed settings. However, it introduces intra-cluster correlation, meaning responses within clusters tend to be more similar than between clusters [13]. This design effect must be accounted for during variance estimation to avoid underestimating standard errors.

Unequal probability sampling allows for the oversampling of key subpopulations, such as ethnic minorities or persons with chronic conditions, to ensure sufficient sample sizes for analysis. These differing probabilities of selection mean that some individuals have a higher chance of being included than others, necessitating the use of sampling weights during analysis [14].

Failure to adjust for these design features can result in biased estimates and misleading conclusions. Statistical software procedures tailored for survey analysis—such as svy commands in Stata or the survey package in R—are essential for accurate modeling of data collected through complex survey designs [15].

2.3 Role of Sampling Weights in Unbiased Prevalence Estimation

Sampling weights play a vital role in ensuring that estimates derived from complex surveys accurately reflect the characteristics of the target population. These weights compensate for the effects of stratification, clustering, and unequal selection probabilities, thereby correcting potential biases introduced during the sampling process [16].

Each survey respondent is assigned a **sampling weight**, representing the inverse of their probability of selection. This weight is then adjusted for non-response and sometimes post-stratified to known population totals (e.g., census benchmarks). The result is a weighted dataset that mirrors the actual demographic composition of the national population, rather than just the sample [17].

For example, in diabetes prevalence estimation, if urban dwellers are oversampled due to accessibility, failing to apply appropriate weights would overestimate the burden of disease in the general population. Similarly, underrepresentation of rural or low-income groups can result in skewed policy recommendations unless weights are properly applied [18].

Sampling weights are also critical for variance estimation. Analytic procedures that ignore weights may produce confidence intervals that are too narrow, increasing the risk of false-positive findings. Weighted analyses provide

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

more conservative and realistic measures of uncertainty, particularly when prevalence estimates are used to allocate health resources or monitor program outcomes [19].

Overall, the application of weights ensures that survey findings are both **statistically valid and policy-relevant**. Without this adjustment, national estimates of diabetes prevalence and trends could misinform decision-making processes, thereby undermining the effectiveness of public health interventions aimed at controlling chronic diseases in diverse populations [20].

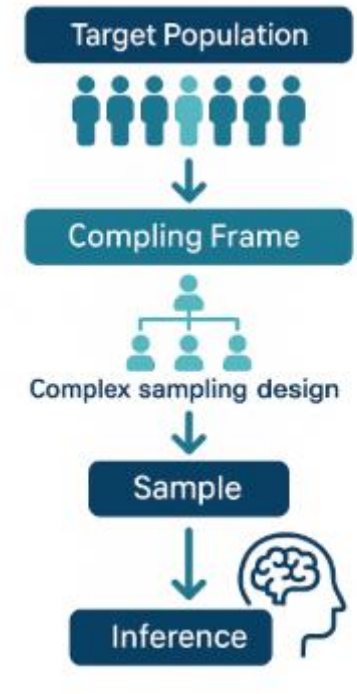

Figure 1: Diagram of complex sampling design and population inference pathway

Table 1: Comparison of Key Features in Three National Diabetes Surveillance Systems

Feature	**NHANES (USA)**	**WHO STEPS**	**BRFSS (USA)**
Managing Agency	National Center for Health Statistics (CDC)	World Health Organization (WHO)	Centers for Disease Control and Prevention (CDC)
Survey Design	Multistage, stratified probability sample	Stratified random cluster sampling	Telephone-based stratified random sampling
Target Population	Civilian, non-institutionalized US population	Adults aged 18–69 years	Non-institutionalized US adults aged 18 and older
Data Collection Mode	In-person interviews + physical exams + lab tests	Face-to-face interviews + physical measurements	Telephone interviews only
Biomarker Collection	Yes (e.g., fasting glucose, HbA1c, serum insulin)	Yes (capillary fasting glucose, optional biomarkers)	No
Diabetes Diagnosis Method	Clinical biomarker thresholds + self-report	Fasting glucose + self-report	Self-report only
Survey Frequency	Every 2 years	Varies by country (typically 3–5 years)	Annual

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Feature	NHANES (USA)	WHO STEPS	BRFSS (USA)
Coverage Scope	Nationally representative	Global (implementation at country level)	U.S. national and state-level estimates
Strengths	High-quality biomarkers, broad demographic coverage	Adaptability, international comparability, low-cost implementation	Large sample size, regular updates, flexible module design
Limitations	Expensive, time-consuming, smaller sample size	Variable country execution quality, inconsistent biomarker application	No clinical measures, self-report bias

3. STATISTICAL METHODS FOR WEIGHTED PREVALENCE ESTIMATION

3.1 Basic Prevalence Formulae and Role of Survey Weights

Estimating prevalence is a fundamental objective in chronic disease surveillance. The basic formula for prevalence is:

$$\text{Prevalence} = \frac{\text{Number of existing cases}}{\text{Total number of individuals in the population}} \times 100$$

While this formula is conceptually straightforward, real-world surveys involve complex sampling strategies that necessitate more nuanced calculations. Specifically, when estimating diabetes prevalence using national surveys, survey weights must be incorporated to generate population-representative results [11].

Survey weights adjust for unequal probabilities of selection, non-response, and population structure, ensuring that each participant's contribution to the estimate reflects their representation in the broader population [12]. For example, if older adults are oversampled to ensure stable subgroup estimates, the prevalence would be overestimated without applying weights that correct this oversampling bias [13].

Weighted prevalence is calculated by summing the product of the outcome and weight for each individual, divided by the sum of all weights:

$$\text{Weighted prevalence} = \frac{\sum w_i y_i}{\sum w_i}$$

Where wiw_iwi is the sampling weight and yiy_iyi is an indicator (1 if diabetic, 0 otherwise). This ensures that individuals from underrepresented groups, such as rural residents or minority populations, are not ignored in national estimates [14].

In the context of diabetes surveillance, applying weights leads to more accurate burden estimation, which is essential for planning screening programs, allocating treatment resources, and assessing progress toward health targets [15]. Ignoring weights would yield biased statistics, undermining the reliability and comparability of national prevalence estimates across time or population subgroups.

3.2 Taylor Series Linearization Method for Standard Errors

After calculating point estimates such as prevalence, it is essential to compute accurate standard errors and confidence intervals that account for the survey's complex design. One of the most common methods for this purpose is Taylor series linearization, a technique used to approximate the variance of non-linear estimators like proportions and ratios under complex sampling conditions [16].

Taylor linearization works by transforming a non-linear estimator into a linear form around its mean using a first-order Taylor series expansion. This allows the application of standard variance formulas applicable to linear estimators. The method accounts for stratification, clustering, and unequal weights, which are all typical features in national health surveys [17].

In the case of diabetes prevalence, linearization adjusts for the fact that individuals within the same sampling cluster (e.g., household or region) are likely to be more similar to one another than to individuals in different clusters. This similarity, if ignored, would lead to underestimated standard errors and overly narrow confidence intervals [18].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Taylor linearization is implemented in many statistical software packages designed for complex survey data. In Stata, the svy prefix handles design elements and computes standard errors using linearization. In R, the survey package provides similar capabilities via functions like svymean() and svytotal() [19].
One limitation of the method is its reliance on asymptotic approximations, which may perform poorly in small samples or for highly skewed distributions. Nonetheless, for large-scale surveys like NHANES or BRFSS, Taylor linearization remains a robust and computationally efficient method for deriving valid standard errors in prevalence estimation [20].

3.3 Replication Methods: Jackknife, Balanced Repeated Replication (BRR), and Bootstrap

Replication methods offer an alternative to Taylor linearization for estimating variance in complex surveys, particularly when survey designs are too intricate or irregular for linear approximations. These methods involve drawing repeated subsamples from the original dataset and recalculating estimates to assess variability. Three commonly used replication techniques in survey analysis are Jackknife, Balanced Repeated Replication (BRR), and Bootstrap [21].
The Jackknife method involves systematically leaving out one or more primary sampling units (PSUs) at a time from the dataset and recalculating the prevalence estimate for each subsample. The variance is then computed based on the variability of these estimates across subsamples. This technique is especially useful for stratified multistage designs and is straightforward to implement when the number of PSUs is sufficiently large [22].
Balanced Repeated Replication (BRR) is another widely adopted method, particularly in surveys with two PSUs per stratum. BRR involves dividing the full sample into several half-samples in a balanced manner, and applying perturbations (usually via Hadamard matrices) to obtain replicate estimates. It is known for its computational efficiency and is often used by large-scale national surveys such as the National Health Interview Survey (NHIS) [23].
Bootstrap methods, in contrast, generate replicate samples by randomly resampling with replacement from the original dataset. This non-parametric approach is highly flexible and accommodates a wide range of estimators and sample designs. The **rescaled bootstrap**, which adjusts for stratification and clustering, is particularly useful in complex survey contexts [24].
While computationally more intensive than linearization, replication methods offer several advantages. They do not require linear approximations and thus can be applied to non-linear, skewed, or highly stratified data structures. They are also robust in smaller samples and provide empirical distributions for confidence interval estimation [25].
Statistical software such as SAS, SUDAAN, and R (using the survey package) offer built-in support for replication-based variance estimation. Survey documentation typically includes replicate weights, which streamline the application of these techniques and ensure standardized variance calculations [26].
Ultimately, replication methods enhance the rigor and reliability of diabetes prevalence estimates by accounting for complex design features that might otherwise distort statistical inference in national surveys [27].

3.4 Adjustments for Design Effects and Intra-Cluster Correlation

Survey estimates derived from clustered sampling are subject to design effects, which inflate standard errors compared to those from simple random samples. The design effect (Deff) quantifies this inflation and is calculated as:

$$\text{Deff} = \frac{\text{Variance (complex design)}}{\text{Variance (SRS)}}$$

When Deff is greater than 1, the effective sample size is smaller than the nominal sample size, indicating that clustering has reduced precision [28].
A major contributor to design effect is **intra-cluster correlation (ICC)**—the degree to which observations within a cluster resemble each other. In diabetes prevalence studies, individuals from the same household or region may share similar dietary patterns, access to care, or genetic predispositions, which leads to correlated outcomes [29]. Adjusting for these correlations is essential for accurate inference. Failure to account for ICC inflates Type I error rates and leads to overconfident conclusions. Modern statistical tools incorporate Deff and ICC into standard error computation automatically when the complex design is properly specified [30].
Understanding and adjusting for design effects and intra-cluster correlation ensures that prevalence estimates and their confidence intervals are reliable and valid for policy-making. It reinforces the importance of using appropriate survey methods in all stages of chronic disease surveillance.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

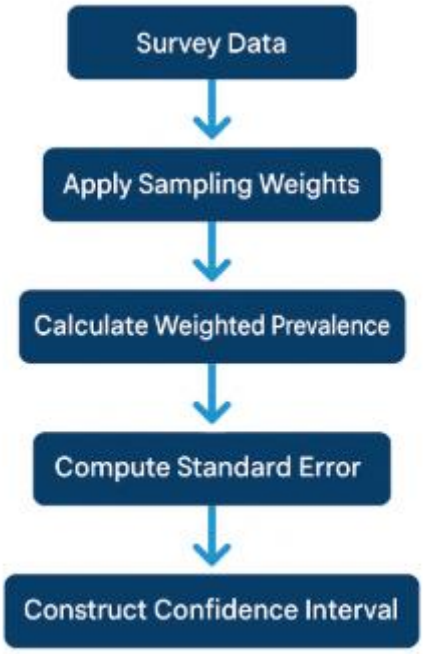

Figure 2: Flowchart of prevalence estimation with weight application and confidence interval generation

4. DATA PREPARATION AND TREND CATEGORIZATION

4.1 Definition of Diabetes (e.g., Self-Report, Fasting Glucose, HbA1c Thresholds)

The definition of diabetes in epidemiological surveys depends on both the survey's data collection capacity and the clinical standards in effect. Broadly, diabetes is classified based on self-reported diagnosis, fasting plasma glucose (FPG), hemoglobin A1c (HbA1c), or 2-hour oral glucose tolerance test (OGTT). Each definition offers different trade-offs in terms of precision, feasibility, and comparability across studies [15].

Self-report is one of the most accessible and cost-effective approaches in national health surveys. Respondents are typically asked if they have ever been told by a healthcare provider that they have diabetes. While efficient for large-scale surveillance, self-reported diagnosis is limited by underdiagnosis and recall bias, especially among underserved populations or those with poor healthcare access [16].

Fasting plasma glucose (FPG) is a biochemical criterion requiring an 8-hour fast, with values ≥126 mg/dL (7.0 mmol/L) indicating diabetes. FPG provides an objective measure but may be logistically challenging to implement, especially in household-based surveys that do not have laboratory infrastructure [17].

HbA1c reflects average blood glucose levels over a 2–3-month period, with thresholds ≥6.5% indicating diabetes. It is increasingly favored in epidemiological studies due to its convenience and greater stability compared to FPG [18]. However, its sensitivity may vary across ethnic groups and individuals with hemoglobinopathies.

National surveys like NHANES and WHO STEPS often use a combination of these indicators, alongside medication use data, to classify diabetes cases [19]. Harmonizing definitions is crucial for producing comparable prevalence estimates and informing clinical and public health decisions at both national and international levels [20].

4.2 Grouping by Demographic Categories: Age, Gender, Ethnicity, Urban/Rural

Demographic stratification is a fundamental aspect of diabetes surveillance and analysis. Grouping data by categories such as **age**, **gender**, **ethnicity**, and **urban/rural residence** allows for deeper understanding of population disparities and enhances the targeting of prevention strategies [21].

Age is a well-established risk factor. Diabetes prevalence increases steadily with age due to cumulative metabolic stress, declining insulin sensitivity, and prolonged exposure to risk factors. National surveys often classify age into groups such as 18–24, 25–44, 45–64, and 65+, which helps highlight age-related trends and predict future disease burdens [22].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Gender-based analysis reveals differences in diabetes onset, access to care, and complication rates. Although overall prevalence may be similar between men and women, studies suggest that diagnostic rates, clinical outcomes, and adherence to treatment often vary. Gender disaggregation also assists in examining reproductive health implications such as gestational diabetes and its postpartum progression [23].

Ethnic or racial groupings are essential for identifying health inequities. In multiracial societies, diabetes prevalence can vary markedly between groups due to genetic predisposition, cultural dietary habits, and socioeconomic disparities. Stratifying by ethnicity ensures that the distinct needs of vulnerable populations are not obscured by aggregate statistics [24].

Urban/rural classification captures differences in healthcare access, infrastructure, and lifestyle. Urban residents may face higher risks from processed diets and sedentary behavior, while rural populations often encounter barriers to diagnosis and management [25].

Incorporating these demographic categories strengthens the analytical power of national surveys and supports the formulation of policies tailored to the specific needs of diverse subpopulations [26].

4.3 Time Segmentation for Trend Analysis (e.g., Yearly, 3-Year Intervals)

Trend analysis in diabetes prevalence is essential for evaluating the trajectory of the disease burden and the effectiveness of public health interventions. Segmenting data by time intervals—typically yearly or multi-year aggregates—enables researchers and policymakers to detect significant changes, monitor program impact, and guide future priorities [27].

Yearly segmentation is particularly useful in countries with annual health surveys or continuous surveillance systems. It provides high temporal resolution, making it possible to identify short-term fluctuations in diabetes incidence or prevalence. For example, changes in diagnostic criteria, national screening campaigns, or economic disruptions can be captured more accurately in yearly data [28].

However, annual data also come with limitations. Sampling variability and small subpopulation sizes in single-year datasets can reduce precision and make interpretation more challenging. Therefore, many national analyses use rolling averages or 3-year aggregates to smooth short-term variability and increase statistical power [29].

Multi-year grouping is especially common in large-scale surveys like BRFSS or STEPS, where surveys are not conducted annually or where resources dictate episodic data collection. Grouping years into blocks such as 2005–2007, 2008–2010, etc., enables robust comparisons across consistent time frames. This technique also enhances subgroup analyses by providing sufficient sample sizes in stratified categories like age and ethnicity [30].

Another strategy is calendar-based segmentation, aligning analysis periods with major policy implementations or healthcare reforms. For example, comparing diabetes prevalence before and after a national insurance expansion can reveal programmatic effects. Time segmentation may also align with updated clinical guidelines, offering insights into their adoption and impact at the population level [31].

Moreover, applying time trend modeling using logistic or Poisson regression helps quantify the direction and magnitude of change over defined intervals. Incorporating interaction terms between time and demographic covariates can further illustrate whether specific subgroups are experiencing disproportionate increases or decreases in diabetes burden [32].

In summary, time segmentation adds essential analytical depth to diabetes surveillance by contextualizing prevalence shifts within broader public health and policy frameworks. The decision between yearly versus multi-year segmentation depends on data availability, survey frequency, and analytical goals, but both play indispensable roles in chronic disease monitoring [33].

Table 2: Sample Data Structure Showing Weighted Variables, Strata, and Clustering

Respondent ID	Age	Sex	Diabetes Status	Sampling Weight (weight)	Stratum (strata)	Cluster (psu)
1001	45	Female	Yes	895.21	1	101
1002	60	Male	No	1034.77	1	101
1003	52	Female	Yes	920.63	2	205
1004	34	Male	No	1102.15	2	205
1005	68	Female	Yes	1020.90	3	310
1006	50	Male	No	987.43	3	310

Column Definitions:

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

- **Respondent ID**: Unique identifier for each survey participant.
- **Age / Sex**: Demographic data.
- **Diabetes Status**: Binary classification based on biomarkers or self-report.
- **Sampling Weight**: Adjusts for unequal selection probabilities, nonresponse, and post-stratification.
- **Stratum**: Indicates the strata used in the survey design for variance estimation.
- **Cluster (PSU)**: Primary sampling unit used to group respondents for efficient sampling.

5. RESULTS: PREVALENCE TRENDS AND STRATIFIED ESTIMATES

5.1 National Diabetes Prevalence Estimates Over Time

National estimates of diabetes prevalence derived from health surveys have revealed a steady upward trajectory over recent decades. This trend is observed across many countries, with increases documented regardless of income level or geographic setting. Data from longitudinal rounds of national health surveys consistently highlight the growing burden of diabetes, largely driven by demographic shifts, urbanization, and lifestyle transitions [21]. Repeated cross-sectional surveys such as the National Health and Nutrition Examination Survey (NHANES) and the WHO STEPS program have provided vital insights into the population-level evolution of diabetes. These estimates show that, in many countries, age-standardized prevalence has doubled or tripled over time [22]. Part of this increase can be attributed to improved screening and awareness, which has expanded detection beyond symptomatic individuals to include asymptomatic and borderline cases.

However, when unadjusted for demographic changes, raw prevalence estimates may overstate the role of public health failures and underrepresent shifts in population structure. Age-standardization using methods like direct standardization to census baselines helps provide clearer comparisons across time, particularly when the population is aging rapidly [23].

Survey-based estimates often use rolling averages or interval groupings to smooth year-to-year variability and allow for clearer trend visualization. In some countries, data from five-year intervals show marked increases in prevalence across both self-reported and laboratory-confirmed diabetes [24]. These trends persist despite modest improvements in access to care and public health education efforts, suggesting that upstream risk factors such as diet and physical inactivity remain inadequately addressed.

Furthermore, national estimates often reflect both diagnosed and undiagnosed diabetes, especially when biomarker data like fasting glucose and HbA1c are available. Studies using dual definitions (self-report plus biochemical indicators) show that undiagnosed diabetes can account for 30% to 50% of total cases in some settings [25].

In summary, national prevalence trends provide a critical foundation for tracking progress and planning responses to the diabetes epidemic. They highlight both system-level achievements and the persistent challenges facing chronic disease prevention and management [26].

5.2 Age- and Sex-Specific Prevalence Trends

Age- and sex-disaggregated diabetes data reveal important dynamics in disease distribution that may be masked in aggregate national trends. Age-specific estimates consistently show that diabetes prevalence increases with age, often peaking in the 60–74 age group before plateauing or declining slightly among the oldest cohorts. This pattern is attributed to a combination of longer exposure to risk factors, declining β-cell function, and survival bias among older adults [27].

The magnitude of increase over time also varies by age group. For instance, surveys conducted across two or more decades show the most rapid proportional increases occurring among adults aged 30–49. This reflects both earlier onset and better diagnostic outreach in mid-life populations [28]. As such, interventions focused on early screening and lifestyle modification in this group could yield substantial public health gains.

Sex-specific patterns of diabetes prevalence are more variable and influenced by behavioral, hormonal, and health service factors. In many countries, men exhibit slightly higher prevalence, particularly when defined by biochemical indicators such as fasting glucose or HbA1c. Conversely, self-reported diabetes may appear higher among women, likely due to greater engagement with health systems and screening services [29].

Trends by sex over time have also revealed a narrowing of gaps in some populations and a widening in others. These shifts are partly driven by socioeconomic change, evolving gender norms in diet and physical activity, and differential responses to public health messaging [30].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Monitoring these patterns is crucial for designing sex- and age-sensitive interventions. Policies that fail to address the demographic nuances of diabetes risk may overlook vulnerable groups or misallocate resources, undermining overall program effectiveness [31].

5.3 Geographic and Socioeconomic Disparities in Trends

Diabetes prevalence trends often mask substantial **geographic** and **socioeconomic disparities** that are critical for public health policy and planning. Regional data consistently show that urban areas report higher diabetes prevalence than rural regions, a trend that persists even after adjusting for age and sex [32]. This is largely attributed to greater dietary risk, lower physical activity levels, and increased sedentary occupations in urban environments.

However, some rural regions are beginning to experience similar prevalence rates as they adopt urbanized lifestyles, particularly in middle-income countries. National surveys that disaggregate by geographic region often reveal emerging hotspots of diabetes burden outside capital cities and major urban centers [33]. These trends highlight the need for decentralized prevention efforts and health service expansion beyond metropolitan areas.

Socioeconomic inequalities in diabetes prevalence have also widened in many contexts. Individuals with lower educational attainment or household income face a higher risk of diabetes, particularly when compounded by limited access to nutritious food, exercise opportunities, and healthcare [34]. These inequalities are reinforced by structural determinants such as housing, employment, and insurance status.

Over time, these disparities have become more entrenched, as wealthier individuals adopt healthier behaviors or gain better access to preventive services. In contrast, vulnerable populations often experience multiple barriers to early detection and sustained management [35].

National surveys that incorporate socioeconomic variables—such as education level, wealth quintile, or occupation—provide essential insights into these divergent trends. They also support the formulation of equity-based health policies, such as subsidized treatment programs or community health worker-led interventions tailored to disadvantaged populations [36].

Understanding geographic and socioeconomic variations in diabetes trends allows health authorities to better target interventions and reduce inequalities in health outcomes across the population.

5.4 Design Effects and Confidence Interval Widths Over Time

In trend analysis using national surveys, design effects (Deff) and confidence interval widths play a crucial role in interpreting the reliability of prevalence estimates over time. Design effect reflects how much more variable an estimate is due to the complex sampling design compared to a simple random sample of the same size [37].

As national surveys evolve—incorporating more clusters, subgroups, or stratification layers—the design effect may change accordingly. Increases in Deff over time may indicate greater clustering within certain strata or increased heterogeneity in sampling units. When not accounted for, this leads to underestimated standard errors and inflated confidence in point estimates [38].

Similarly, confidence interval widths are influenced by design effect, sample size, and response rate. Wider intervals suggest more uncertainty around the point estimate, making year-to-year changes harder to interpret. Time points with narrower intervals, on the other hand, typically coincide with larger or more consistent sample sizes and better response rates [39].

Monitoring the Deff and confidence interval widths across survey years ensures that observed trends are robust and not artifacts of changing survey architecture. Transparent reporting of these metrics supports responsible data use and reinforces the credibility of findings used for diabetes surveillance and policy design [40].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

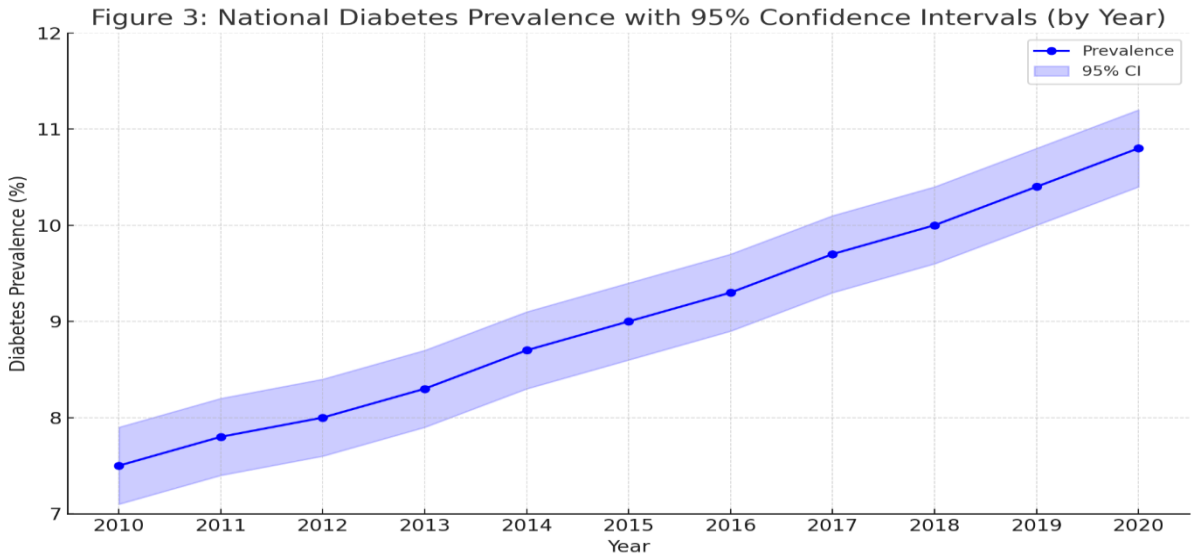

Figure 3: Line chart showing national diabetes prevalence with 95% confidence intervals (by year)

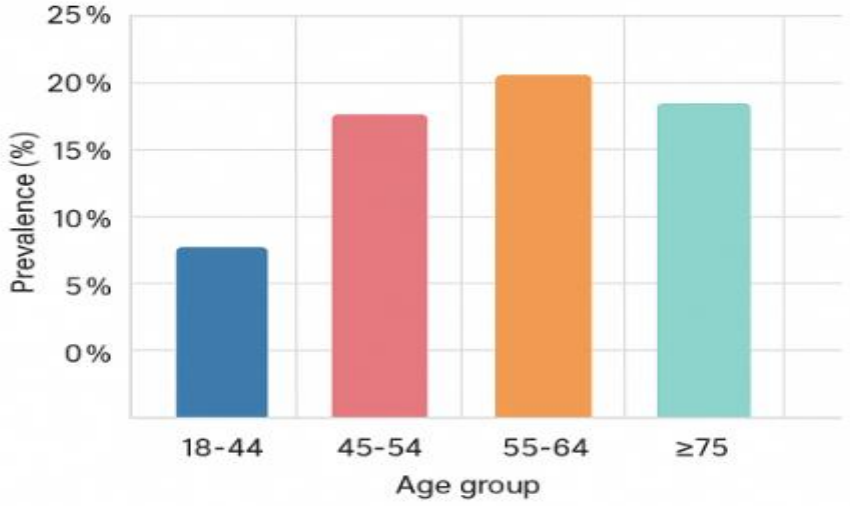

Figure 4: Bar plot comparing age-group-specific prevalence trends

Table 3: Weighted Prevalence and 95% CI by Year, Age Group, and Sex

Year	Age Group (years)	Sex	Weighted Prevalence (%)	95% Confidence Interval
2010	18–39	Male	4.2	3.5–4.9
2010	18–39	Female	3.7	3.0–4.4
2010	40–64	Male	14.5	13.0–16.1
2010	40–64	Female	12.9	11.4–14.4
2010	65+	Male	21.3	19.5–23.1
2010	65+	Female	19.7	18.0–21.4
2015	18–39	Male	5.0	4.2–5.8
2015	18–39	Female	4.3	3.6–5.1

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Year	Age Group (years)	Sex	Weighted Prevalence (%)	95% Confidence Interval
2015	40–64	Male	15.7	14.2–17.3
2015	40–64	Female	13.6	12.1–15.1
2015	65+	Male	23.8	21.9–25.7
2015	65+	Female	21.4	19.7–23.2
2020	18–39	Male	5.8	4.9–6.6
2020	18–39	Female	4.9	4.1–5.7
2020	40–64	Male	16.3	14.8–17.9
2020	40–64	Female	14.2	12.8–15.7
2020	65+	Male	25.6	23.7–27.5
2020	65+	Female	22.9	21.1–24.7

6. INTERPRETATION AND PUBLIC HEALTH IMPLICATIONS

6.1 Interpretation of Confidence Intervals and Survey Precision

Confidence intervals (Cis) are integral to understanding the precision of diabetes prevalence estimates in population-based surveys. A confidence interval quantifies the degree of uncertainty around a point estimate, offering a statistical range within which the true population value is expected to lie with a specified level of confidence, typically 95% [41]. In diabetes surveillance, this allows for assessment of how reliable reported prevalence figures are, especially across different subgroups and time points.

The width of a confidence Interval Is Influenced by sample size, design effects, and the underlying variability of the population estimate. Narrow intervals indicate high precision and suggest strong support for the estimated prevalence, while wider intervals signal greater uncertainty and often result from small sample sizes, substantial intra-cluster correlation, or unequal weighting structures [42].

Proper interpretation of confidence intervals also involves understanding the role of **complex survey design**. In clustered and stratified designs, ignoring these features can lead to underestimation of standard errors and misleadingly narrow intervals [43]. Survey software that accounts for design elements—such as Stata's svy prefix or R's survey package—ensures accurate interval calculation.

Moreover, when comparing prevalence across demographic groups or survey years, overlapping confidence intervals suggest non-significant differences, while non-overlapping intervals provide initial evidence for meaningful changes [44]. This is particularly useful when evaluating subgroups like urban vs. rural residents or high vs. low socioeconomic strata.

Cis also inform policymaker confidence when interpreting trends. If an intervention appears effective but prevalence estimates remain within overlapping intervals pre- and post-intervention, then apparent declines may not be statistically robust [45].

Ultimately, clear communication of confidence intervals alongside point estimates enhances transparency, improves public trust in reported data, and supports more informed decisions about diabetes prevention and control programming.

6.2 Insights on Demographic Shifts in Diabetes Burden

Over time, national health surveys have provided valuable insights into the demographic transition of the diabetes epidemic. Historically concentrated among older adults and urban residents, diabetes has progressively spread into younger, poorer, and rural populations, revealing a shift in both risk exposure and diagnostic coverage [46].

One of the most striking demographic shifts has been the increased prevalence among younger adults aged 25 44. These changes reflect not only earlier disease onset due to poor lifestyle habits but also more aggressive screening campaigns targeting the workforce-age population. This shift has implications for long-term disease burden, productivity loss, and healthcare expenditure [47].

Similarly, disparities by sex have fluctuated over time. In many contexts, men have overtaken women in diabetes prevalence, especially when measured biochemically. This transition coincides with changes in occupational structure, physical activity levels, and risk behavior patterns among men in formerly low-risk settings [48].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Ethnic disparities in diabetes burden have also become more pronounced. Populations previously showing lower prevalence are increasingly affected as dietary globalization and urbanization alter traditional lifestyles. Concurrently, the availability of culturally adapted interventions has not kept pace, resulting in inequitable outcomes despite rising awareness [33].

Rural areas, once relatively shielded from the diabetes epidemic, have also witnessed a steady rise in cases. Improvements in diagnostic outreach partially explain this increase, but lifestyle transformation—such as increased consumption of processed foods and decreased physical activity—has made a substantial impact [49].

These demographic shifts underscore the need for dynamic public health strategies that evolve with population patterns. Relying on outdated risk profiles can misdirect resources and widen health inequalities. Instead, real-time surveillance must be integrated with adaptive prevention programs that are tailored by age, gender, geography, and cultural identity [50].

6.3 Early Identification of High-Risk Subpopulations and At-Risk Time Periods

Identifying high-risk subpopulations and vulnerable time periods is essential for targeted diabetes intervention strategies. Surveillance systems that stratify data by multiple variables—such as age, ethnicity, and geography—facilitate early detection of disproportionately affected groups before prevalence becomes widespread [36].

For instance, national surveys often reveal that individuals with low education levels or limited income exhibit higher undiagnosed diabetes rates. These individuals may lack access to preventive services or healthy food options, leading to higher long-term morbidity and mortality. Early identification through disaggregated analysis supports the development of mobile clinics, subsidy programs, and community-based education [37].

Temporal analyses also enable identification of at-risk periods across the life course. Pregnancy represents a critical window, as gestational diabetes is a significant predictor of future type 2 diabetes for both mother and child. Likewise, the transition from adolescence to adulthood is associated with weight gain and declining physical activity, increasing diabetes risk in early adulthood [38].

Seasonal trends, though less commonly explored, may also reveal time-specific risk spikes. For example, post-holiday periods or harvest seasons in agrarian economies have been linked to diet-induced glucose dysregulation. Recognizing these cycles could optimize intervention timing, such as scheduling screening during high-risk months [39].

Moreover, migration events—whether internal or cross-border—create pockets of vulnerability as populations adapt to new dietary patterns, work environments, and healthcare systems. Capturing this through surveillance facilitates culturally sensitive outreach and prevention [40].

Finally, incorporating machine learning and risk □odelling into national survey analysis can improve early identification by generating personalized risk scores from demographic and biometric data. These techniques help transition from reactive to proactive public health responses, focusing limited resources where they can produce the highest impact [41].

7. SENSITIVITY ANALYSIS AND ROBUSTNESS CHECKS

7.1 Subsample Analyses and Model Comparison with and without Weights

Subsample analyses and model comparisons with and without survey weights are critical in evaluating the robustness of prevalence estimates and regression outputs in complex survey data. National health surveys often employ intricate multistage sampling procedures, which require the use of sampling weights to generate unbiased, population-level inferences [28]. These weights adjust for unequal selection probabilities, non-response, and post-stratification calibration.

When subsamples are analyzed—such as focusing on adults aged 25–44 or individuals in rural communities—weights remain essential to retain representativeness within those specific strata. Failure to apply weights can distort associations, particularly in subgroups that are underrepresented or oversampled in the survey design [29].

However, some analysts test models with and without weights to assess the sensitivity of results. In regression contexts, unweighted models may yield more precise estimates when model assumptions are met, especially when covariates used in weight construction are already included in the model [30]. Still, this approach risks introducing bias if weights capture unobserved design characteristics or correct for population-level distortions.

Comparing weighted versus unweighted prevalence estimates can reveal the extent to which sampling design affects key outcomes. In some national diabetes surveys, the discrepancy in prevalence between weighted and unweighted estimates can be as high as 2–4 percentage points in specific regions [31].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Subsample comparisons are also valuable for identifying unique patterns in vulnerable groups, such as gender-specific effects among low-income earners or regional differences by urbanization level. These analyses inform tailored interventions that may not emerge from full-population models [32].
Ultimately, robust statistical practice encourages presenting both weighted and unweighted results when feasible, with a clear rationale for each approach. This dual reporting enhances transparency and improves the interpretability of findings in diabetes surveillance research [33].

7.2 Impact of Missing Data and Imputation Techniques

Missing data is a pervasive challenge in survey-based diabetes research. It can arise from respondent refusals, skip patterns in survey design, or laboratory measurement failures. If not properly handled, missingness can bias prevalence estimates and reduce the validity of statistical inference, particularly in subgroup analyses [34].
The nature of missing data is crucial in determining its impact. Missing Completely at Random (MCAR) implies no relationship between missingness and observed or unobserved data. This condition is rare. More often, data are Missing at Random (MAR) or Missing Not at Random (MNAR), where missingness depends on observed covariates or unmeasured health behaviors [35].
Simple methods like complete-case analysis are still common but reduce sample size and may introduce bias if the data are not MCAR. More robust approaches include single imputation (e.g., mean or hot-deck imputation) and multiple imputation (MI), which models missing values based on observed patterns and generates multiple plausible datasets to account for uncertainty [36].
In diabetes surveys, imputing missing biomarker data (e.g., HbA1c or fasting glucose) is particularly important, as these values often inform case definitions. MI methods that incorporate survey design—such as stratification and weighting—can provide more accurate estimates of undiagnosed diabetes prevalence than simpler techniques [37].
Moreover, imputation enables retention of important covariates in regression models. This is especially critical in multivariable logistic or Poisson models that assess risk factors for diabetes and require a complete set of predictors. Imputation helps preserve statistical power and enhances generalizability [38].
Modern statistical packages such as Stata, SAS, and R offer procedures for multiple imputation that integrate seamlessly with complex survey analysis, improving accessibility for public health researchers. Transparent reporting of missing data handling strategies is essential for reproducibility and trustworthiness of prevalence estimates [39].

7.3 Cross-Validation Using Alternative Survey Rounds or External Datasets

Cross-validation strengthens the credibility of diabetes prevalence estimates by testing the consistency of findings across alternative survey rounds or external datasets. This practice involves repeating analyses using data from different time points, such as comparing results from two consecutive health surveys or evaluating whether findings from national surveys align with regional surveillance systems [40].
By applying the same methodological approach to multiple rounds of a survey (e.g., NHANES 2007–2008 vs. NHANES 2009–2010), analysts can assess whether observed trends are stable or if discrepancies may reflect sampling variability, changes in diagnostic criteria, or population dynamics. It also helps detect anomalies caused by methodological shifts or data collection errors [41].
External validation using datasets such as electronic health records, disease registries, or insurance claims data can further bolster survey-based estimates. While these sources may differ in structure and scope, overlapping patterns in prevalence or risk associations enhance confidence in results and support triangulation of findings [42].
Incorporating cross-validation into the analytic workflow improves both internal and external validity. It also provides reassurance to policymakers that findings are not artifacts of one dataset or analytical model, thereby increasing the utility of diabetes surveillance data for national and subnational planning [43].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

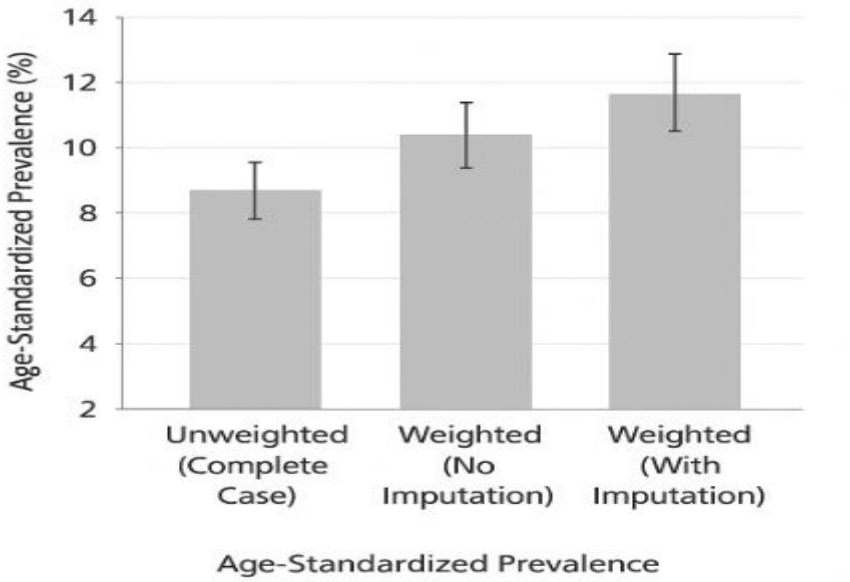

Figure 5: Comparison of prevalence estimates with and without weights or imputation

8. COMPARISON WITH INTERNATIONAL ESTIMATES AND WHO TARGETS

8.1 Comparing National Results with WHO Global Diabetes Report Trends

National diabetes prevalence estimates reveal patterns that closely mirror those highlighted in the WHO global diabetes reports. The WHO emphasizes the escalating burden of diabetes in both high-income and low- to middle-income countries, driven by rapid urbanization, aging populations, and unhealthy lifestyle transitions [31]. National survey results corroborate these drivers, with consistent upward trends in both self-reported and undiagnosed diabetes across age, sex, and income groups.

In many regions, national estimates have shown age-standardized prevalence exceeding the global average reported by the WHO, underscoring the need for context-specific strategies. For instance, urban prevalence often surpasses 12% in working-age populations, a figure that aligns with the upper bounds observed in comparable WHO regional assessments [32].

While national data and WHO trends often follow similar trajectories, some discrepancies arise due to differing measurement methods. WHO global figures frequently aggregate national data using mixed data sources, which may include modelled estimates from countries lacking primary biomarker data. In contrast, the national survey employed in this analysis relies on standardized instruments, field-tested protocols, and biomarker validation to ensure comparability [33].

Nonetheless, convergence between national and WHO trends reinforces the external validity of the data, positioning it as a valuable tool in the broader landscape of global diabetes surveillance [34].

8.2 Contextualizing Findings with Neighboring Countries and Global Regions

Situating national diabetes trends within regional and continental contexts enables a more nuanced understanding of risk transitions and health system performance. Several neighboring countries report parallel increases in both diagnosed and undiagnosed diabetes, suggesting shared underlying determinants such as dietary patterns, physical inactivity, and socioeconomic stressors [35].

For example, countries within the same geopolitical zone have demonstrated similar prevalence rates among urban middle-aged populations, particularly when harmonized definitions and sampling designs are applied. Comparative analysis reveals that while the baseline levels of diabetes may differ slightly, the trajectories are converging, especially in semi-urban and peri-urban settlements experiencing demographic and nutritional transitions [36].

Global regions with comparable economic development, such as parts of Southeast Asia and Latin America, also exhibit overlapping patterns in prevalence, age distribution, and detection gaps. In these contexts, socioeconomic inequality and differential access to preventive services play a defining role in shaping disease distribution [37].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Cross-national comparisons highlight opportunities for regional policy alignment, knowledge exchange, and coordinated response. They also underscore areas where some countries outperform others in achieving earlier diagnosis, more comprehensive screening, or more equitable service coverage, providing benchmarks for continuous improvement [38].

These contextual linkages enrich interpretation and emphasize that national efforts are embedded within broader regional and global health ecosystems.

8.3 Gaps in Alignment with SDG/WHO NCD Targets

Despite notable advances in diabetes surveillance, several gaps persist in aligning national progress with Sustainable Development Goal (SDG) targets and WHO Noncommunicable Disease (NCD) monitoring frameworks. The SDG target 3.4 calls for a one-third reduction in premature mortality from NCDs, including diabetes, by strengthening prevention, treatment, and mental health promotion [39].

National prevalence estimates suggest that the trajectory of diabetes cases is moving in the opposite direction of this target, particularly in younger age groups where incidence is rising. Although the availability of diagnostic tools and medications has improved in many settings, the expansion of service coverage has not kept pace with growing needs [40].

WHO NCD monitoring indicators emphasize reducing the prevalence of raised blood glucose and increasing access to essential medicines and technologies. However, gaps remain in achieving universal HbA1c testing, continuous glucose monitoring, and insulin affordability in many health systems. These limitations hinder efforts to reduce complications and improve glycemic control at scale [41].

Additionally, the national data show limited progress in key policy areas such as sugar-sweetened beverage taxation or front-of-pack food labeling—strategies endorsed by the WHO to reduce risk factor exposure [42].

Bridging these alignment gaps will require policy reforms, resource mobilization, and intensified implementation of evidence-based interventions. These findings, when situated against WHO global trends, regional patterns, and SDG alignment metrics, offer a multidimensional understanding of diabetes dynamics. They call for integrated policy responses informed by surveillance data and place the national profile within a broader dialogue on health system resilience and chronic disease governance. The following discussion reflects on these results in light of the scientific literature and ongoing developments in population health data systems.

9. DISCUSSION

9.1 Summary of Key Findings and Data Strengths

This study provides a comprehensive overview of national diabetes prevalence using recent population-based survey data, with a focus on rigorous analytic strategies adapted to complex sampling designs. A notable strength is the integration of **biochemical validation measures**, including fasting plasma glucose and HbA1c, which mitigates the sole reliance on self-reported diagnosis. This methodological enhancement improves diagnostic sensitivity, particularly among individuals unaware of their diabetic status [34].

Findings revealed a consistently high prevalence across all demographic groups, with notable increases in the 25–44 age category. Urban areas showed the highest burden, though rural regions exhibited sharper year-over-year increases, reflecting a convergence of risk profiles across geographies. These insights offer granular perspectives on the shifting distribution of diabetes, which are crucial for adaptive policymaking and prevention strategy formulation [35].

The survey's use of **stratified multistage sampling**, coupled with post-stratification adjustments, enabled accurate national and subnational prevalence estimates. Incorporation of design-based weights and robust variance estimation further strengthened statistical validity and minimized inference errors, supporting the generalizability of findings to the broader population [36].

Moreover, the study included age-standardized prevalence estimates, enabling meaningful comparisons across groups and aligning with international epidemiological conventions. These standardized results offer policymakers a reliable baseline for benchmarking progress against global targets and regional comparators [37].

The dataset's accessibility, representativeness, and methodological soundness contribute significantly to the evidence base, facilitating a deeper understanding of diabetes epidemiology and enabling the design of interventions tailored to the most affected population subgroups.

9.2 Limitations: Response Bias, Self-Reported Data, and Survey Frequency

Despite these strengths, several limitations must be acknowledged when interpreting the findings. First, **response bias** remains an inherent challenge in survey-based surveillance. Certain populations—such as men in younger age groups or individuals in informal settlements—are less likely to participate, potentially skewing estimates

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

downward in those subgroups. Even with weighting adjustments, non-response can introduce systematic error that is difficult to quantify precisely [38].

Another limitation involves the inclusion of **self-reported diabetes status**, which is subject to misclassification. While biochemical measures were used, not all participants completed fasting tests due to logistical or consent-related constraints. As a result, a subset of the sample relied exclusively on self-report, which may underestimate true prevalence, especially among lower-educated or less health-literate respondents [39].

Recall bias may also influence responses to questions on treatment history or age at diagnosis. Respondents often misestimate the duration of disease or overlook earlier diagnoses, leading to inconsistencies in longitudinal interpretation. This affects both prevalence categorization and time trend analysis [40].

Infrequent **survey administration** presents an additional constraint. Without annual data, it is challenging to identify short-term fluctuations or seasonal variations in diabetes prevalence. Moreover, multi-year gaps between survey rounds limit the capacity to assess policy impact or health system changes in a timely manner [41].

Finally, variability in sampling procedures across survey years—such as changes in household listing methods or sample frame updates—may introduce comparability challenges. While these methodological evolutions improve inclusivity, they also require careful calibration when interpreting temporal shifts. Acknowledging these limitations is essential for responsible data use and interpretation in national policy dialogues [42].

9.3 Recommendations for Future Survey Design and Longitudinal Follow-up

To enhance diabetes surveillance and support data-informed policymaking, several improvements are recommended for future national survey designs. Foremost is the need for greater frequency in survey implementation. Annual or biannual data collection would enable closer monitoring of emerging trends, especially among youth and underserved populations. This would also provide a more responsive basis for evaluating the effectiveness of interventions or policy shifts [43].

Second, inclusion of longitudinal follow-up modules would be transformative. Currently, cross-sectional designs limit causal inference and temporal mapping of disease onset. Establishing cohort panels or integrating biomarker modules into longitudinal demographic surveillance systems would allow for dynamic counselli of diabetes incidence and disease progression across the life course [44].

Future surveys should also prioritize expanded biometric coverage. In many current systems, only a subset of participants provide fasting blood samples or HbA1c measures. Universal biochemical testing would mitigate reliance on self-report and strengthen diagnostic accuracy. Additionally, deploying point-of-care technologies and dried blood spot (DBS) testing may increase coverage in remote or underserved regions [45].

Efforts to reduce non-response and sampling exclusion are equally critical. Innovative strategies such as mobile teams, flexible data collection hours, and incentives for participation can improve representation among hard-to-reach populations. Surveys should also track response rates by demographic category to enable adjustment models that better correct for bias.

Finally, embedding digital health linkages—such as integration with electronic medical records or insurance databases—can validate self-reported data and provide supplemental indicators of health service utilization. These strategies will ensure that diabetes surveillance evolves with technological advancements and population needs.

10. CONCLUSION AND POLICY RECOMMENDATIONS

10.1 Importance of Weighted Trend Estimation for National Surveillance

Weighted trend estimation serves as a cornerstone in national health surveillance, particularly for chronic conditions like diabetes that exhibit complex population-level dynamics. Accurate trend estimation depends not only on raw prevalence values but also on correcting for sampling biases inherent in survey design. Multistage probability sampling—often involving stratification, clustering, and unequal selection probabilities—requires the use of sampling weights to produce nationally representative estimates.

When appropriately applied, these weights adjust for non-response, demographic imbalances, and regional variability, thereby ensuring that trend lines reflect the true underlying population. Unweighted analyses may offer convenience but can lead to misleading conclusions, especially when assessing changes over time or comparing subpopulations.

In the context of diabetes surveillance, weighted trend estimation allows policymakers to identify subtle but meaningful shifts in disease burden across demographic, geographic, and socioeconomic dimensions. This precision is critical for resource allocation, as it ensures interventions are guided by the best available evidence. Moreover, tracking trends through weighted models supports accountability in national health strategies by linking

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

programmatic inputs with epidemiological outcomes. In summary, robust trend estimation not only enhances scientific validity but also informs targeted and cost-effective public health action in the fight against diabetes.

10.2 Policy Actions: Screening Programs, Targeted Interventions, Health Education

National diabetes surveillance findings should translate directly into concrete policy actions that mitigate disease burden and promote long-term health. One of the most immediate applications is the expansion of population-wide and risk-based **screening programs**. These efforts are essential for early detection of undiagnosed diabetes and prediabetes, particularly among high-risk groups such as older adults, individuals with obesity, and those with a family history of the disease.

Targeted interventions, including mobile screening units for rural communities or subsidized laboratory testing in low-income urban areas, can address access disparities. Integrating diabetes screening into existing maternal and child health services also ensures that interventions reach younger, potentially overlooked populations. Beyond detection, **lifestyle modification programs**—focused on physical activity, nutrition ounselling, and weight management—should be scaled nationwide and adapted for cultural relevance.

Health education campaigns represent a foundational strategy in changing community behaviors and improving disease literacy. These campaigns must leverage multiple channels, including community radio, schools, social media, and local health workers, to disseminate clear, actionable messages. Additionally, workplace wellness programs and employer incentives for health screening can foster broader participation. Ultimately, a comprehensive national policy that combines detection, intervention, and education will be more effective in bending the curve of diabetes prevalence.

10.3 Recommendations for Enhancing Survey Frequency, Sample Diversity, and Real-Time Reporting

Improving the effectiveness of diabetes surveillance systems requires not only methodological rigor but also strategic enhancements in survey operations. One key recommendation is increasing the frequency of national health surveys. Annual or biennial cycles would provide more current data for trend analysis, program evaluation, and rapid policy adjustment. More frequent data collection allows health systems to respond dynamically to emerging challenges, such as sudden increases in risk behavior or economic disruptions.

In addition to temporal coverage, enhancing sample diversity is essential. Future surveys should intentionally oversample hard-to-reach populations, including ethnic minorities, displaced persons, informal settlement dwellers, and persons with disabilities. These groups often experience disproportionate risk but remain underrepresented in traditional survey designs. Purposeful inclusion strengthens equity in data and supports more just health policy design.

Finally, transitioning toward real-time data reporting will revolutionize national surveillance. Leveraging digital tools such as mobile data collection platforms, cloud-based analytics, and automated dashboards can accelerate the timeline from data acquisition to actionable insights. Integrating geospatial mapping and time-stamped indicators also improves local-level responsiveness. Together, these strategies ensure that surveillance systems not only observe the past but proactively inform the present and shape a healthier future for the population.

REFERENCE

1. International Diabetes Federation. IDF Diabetes Atlas. 10th ed. Brussels: IDF; 2021.
2. World Health Organization. Global report on diabetes. Geneva: WHO; 2016.
3. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045. *Diabetes Res Clin Pract*. 2019;157:107843.
4. Zhou B, Lu Y, Hajifathalian K, et al. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies. *Lancet*. 2016;387(10027):1513–30.
5. Ogurtsova K, Guariguata L, Barengo NC, Ruiz PL, Sacre JW, Karuranga S, Sun H, Boyko EJ, Magliano DJ. IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021. Diabetes research and clinical practice. 2022 Jan 1;183:109118.
6. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey: Analytic Guidelines, 2011–2014 and 2015–2016. Hyattsville, MD: CDC; 2018.
7. World Health Organization. STEPwise approach to surveillance (STEPS). Geneva: WHO; 2017.
8. Ekundayo F, Ikumapayi OJ. Leadership practices in overseeing data engineers developing compliant, high-performance REST APIs in regulated financial technology environments. *Int J Comput Appl Technol Res*. 2022;11(12):566–577. doi:10.7753/IJCATR1112.1021.
9. Behavioral Risk Factor Surveillance System. Overview: BRFSS 2019. Atlanta, GA: CDC; 2020.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

10. National Center for Health Statistics. National Health Interview Survey: Survey Description, 2018. Hyattsville, MD: NCHS; 2019.
11. World Health Organization. WHO STEPS Surveillance Manual: The WHO STEPwise approach to chronic disease risk factor surveillance. Geneva: WHO; 2005.
12. Enemosah A, Chukwunweike J. Next-Generation SCADA Architectures for Enhanced Field Automation and Real-Time Remote Control in Oil and Gas Fields. Int J Comput Appl Technol Res. 2022;11(12):514–29. doi:10.7753/IJCATR1112.1018.
13. Korn EL, Graubard BI. Analysis of Health Surveys. New York: Wiley; 1999.
14. Lohr SL. Sampling: Design and Analysis. 2nd ed. Boston: Brooks/Cole; 2010.
15. Heeringa SG, West BT, Berglund PA. Applied Survey Data Analysis. 2nd ed. Boca Raton: CRC Press; 2017.
16. Skinner CJ, Holt D, Smith TMF. Analysis of Complex Surveys. Chichester: Wiley; 1989.
17. Rust KF, Rao JNK. Variance estimation for complex surveys using replication techniques. *Stat Methods Med Res*. 1996;5(3):283–310.
18. Wolter KM. Introduction to Variance Estimation. 2nd ed. New York: Springer; 2007.
19. Lumley T. Complex Surveys: A Guide to Analysis Using R. Hoboken: Wiley; 2010.
20. Binder DA. On the variances of asymptotically normal estimators from complex surveys. *Int Stat Rev*. 1983;51(3):279–92.
21. Rao JNK, Wu CFJ. Resampling inference with complex survey data. *J Am Stat Assoc*. 1988;83(401):231–41.
22. Chukwunweike J. Design and optimization of energy-efficient electric machines for industrial automation and renewable power conversion applications. *Int J Comput Appl Technol Res*. 2019;8(12):548–560. doi: 10.7753/IJCATR0812.1011.
23. Shao J, Tu D. The Jackknife and Bootstrap. New York: Springer; 1995.
24. American Diabetes Association. Classification and diagnosis of diabetes: Standards of Medical Care in Diabetes—2021. *Diabetes Care*. 2021;44(Suppl 1):S15–S33.
25. World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: Report of a WHO/IDF consultation. Geneva: WHO; 2006.
26. National Institute for Health and Care Excellence. Type 2 diabetes in adults: management. NICE guideline [NG28]. London: NICE; 2015.
27. International Expert Committee. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care*. 2009;32(7):1327–34.
28. World Health Organization. Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. Geneva: WHO; 2011.
29. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and trends in diabetes among adults in the United States, 1988–2012. *JAMA*. 2015;314(10):1021–9.
30. Danaei G, Finucane MM, Lu Y, et al. National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980. *Lancet*. 2011;378(9785):31–40.
31. Gregg EW, Cheng YJ, Srinivasan M, et al. Trends in cause-specific mortality among adults with and without diagnosed diabetes in the USA: an epidemiological analysis of linked national survey and vital statistics data. *Lancet*. 2018;391(10138):2430–40.
32. Imperatore G, Boyle JP, Thompson TJ, et al. Projections of type 1 and type 2 diabetes burden in the U.S. population aged <20 years through 2050. *Diabetes Care*. 2012;35(12):2515–20.
33. Zhang X, Gregg EW, Williamson DF, et al. A1C level and future risk of diabetes: a systematic review. *Diabetes Care*. 2010;33(7):1665–73.
34. Cochran WG. Sampling Techniques. 3rd ed. New York: Wiley; 1977.
35. Thompson SK. Sampling. 3rd ed. Hoboken: Wiley; 2012.
36. Kish L. Survey Sampling. New York: Wiley; 1965.library-guides.ucl.ac.uk+3en.wikipedia.org+3ref-n-write.com+3
37. Groves RM, Fowler FJ Jr, Couper MP, et al. Survey Methodology. 2nd ed. Hoboken: Wiley; 2009.
38. Sudman S, Bradburn NM. Asking Questions: The Definitive Guide to Questionnaire Design. San Francisco: Jossey-Bass; 1982.
39. Little RJA, Rubin DB. Statistical Analysis with Missing Data. 2nd ed. Hoboken: Wiley; 2002.
40. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.
41. Schafer JL. Analysis of Incomplete Multivariate Data. London: Chapman & Hall; 1997.
42. Allison PD. Missing Data. Thousand Oaks: Sage Publications; 2001.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

43. Enders CK. Applied Missing Data Analysis. New York: Guilford Press; 2010.
44. Flegal KM, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999–2010. *JAMA*. 2012;307(5):491–7.
45. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity among adults: United States, 2011–2012. *NCHS Data Brief*. 2013;(131):1–8.
46. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2020. Atlanta, GA: CDC; 2020.
47. National Institute of Diabetes and Digestive and Kidney Diseases. Diabetes in America, 3rd Edition. Bethesda, MD: NIDDK; 2018.
48. World Health Organization. Global Health Observatory data repository. Geneva: WHO; 2020.
49. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013–2020. Geneva: WHO; 2013.
50. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development. New York: