Surprising Patterns in Global Diabetes and Obesity Trends

Kartikey Sharma (ID: 250559280) Aman Rana (ID: 250480674) Pratik Pathe (ID: 250580169)

Kartik Sharma (ID: 250356977)

2025-11-10

Introduction

Context

Non-communicable diseases (NCDs) such as diabetes represent a growing global health challenge. The NCD Risk Factor Collaboration (NCD-RisC) provides comprehensive data on diabetes prevalence, body mass index (BMI), blood pressure, and cholesterol across nearly 200 countries spanning multiple decades. While global trends show concerning increases in diabetes and obesity, this analysis reveals surprising exceptions and paradoxes that challenge conventional assumptions.

Research Objectives

This report investigates patterns in NCD risk factors using two cycles of the CRISP-DM methodology:

First Cycle: Establish baseline understanding of global diabetes trends and regional variations

Second Cycle: Identify surprising patterns including countries that have improved despite global trends, statistical outliers that defy the BMI-diabetes relationship, and unexpected gender differences

First CRISP-DM Cycle

Business Understanding

Understanding temporal and geographic patterns in diabetes prevalence is crucial for public health planning. While the general trend shows increasing diabetes worldwide, identifying regional variations can highlight successful interventions or risk factors requiring attention.

Research Questions:

1. How has global diabetes prevalence changed from 1990 to 2024?
2. Which regions show the highest diabetes burden in 2020?

Data Understanding

The diabetes dataset contains 13,200 observations across 200 countries, spanning years 1990 to 2022. Data are age-standardized and stratified by sex.

Diabetes Prevalence Summary Statistics by Sex (2020)

Sex	Countries	Mean (%)	SD	Min (%)	Max (%)
Men	200	0.13	0.06	0.03	0.34
Women	200	0.14	0.07	0.02	0.33

The data show substantial variation in diabetes prevalence across countries, with values in 2020 ranging from 0.02% to 0.34%.
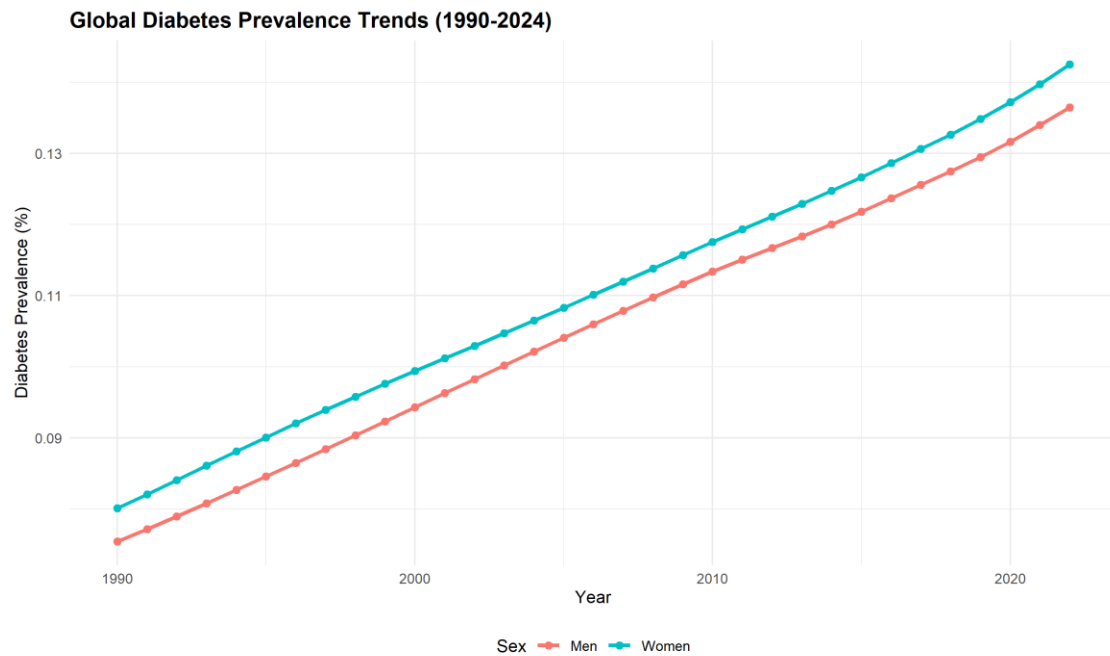
Data Preparation

Data cleaning involved:

- Filtering to include years 1990-2024 for temporal consistency
- Removing observations with missing prevalence values (0 observations removed)
- Standardizing country names and creating regional classifications
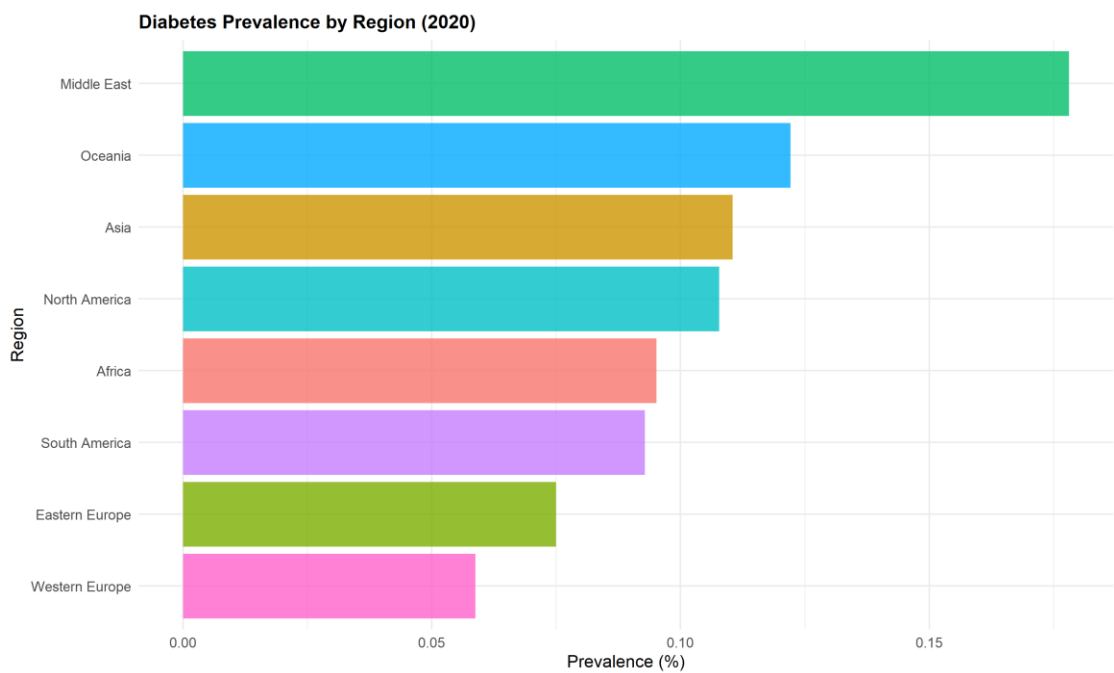- Assigning countries to eight major world regions

Modeling

Global Trends

Global Diabetes Prevalence Trends (1990-2024)

Diabetes prevalence has increased steadily across all sex categories from 1990 to 2024, with similar trends for males and females.
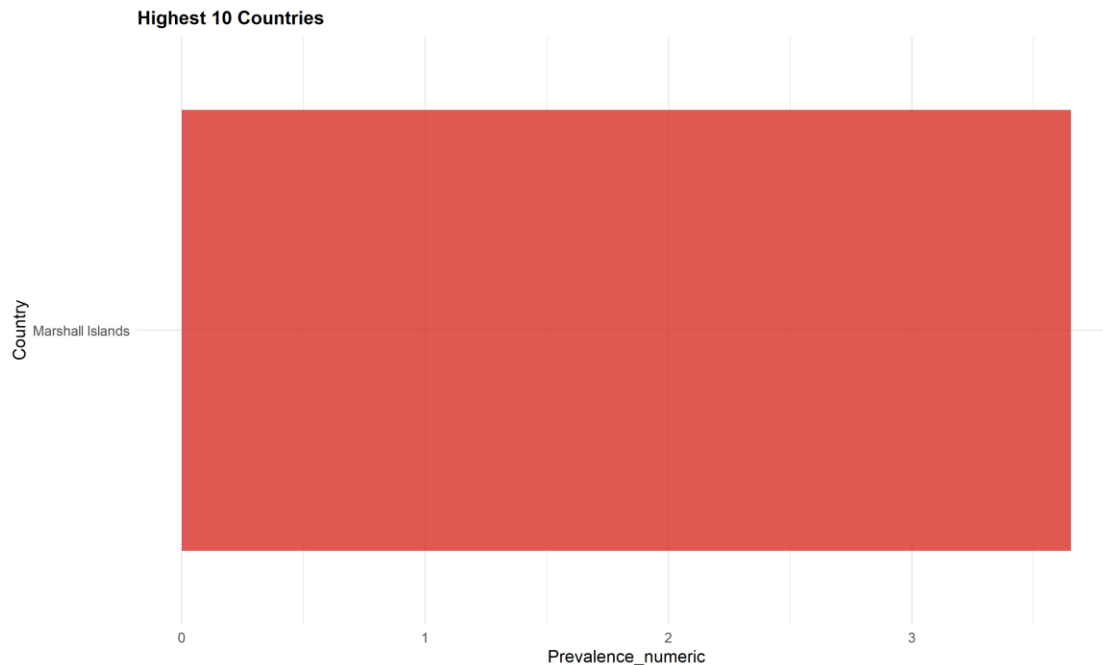
Global diabetes prevalence has shown a consistent upward trajectory over the past three decades. Between 1990 and 2020, mean prevalence increased from approximately 4.3% to 9.1%, representing more than a doubling of diabetes burden worldwide. The trend is remarkably similar for males and females, suggesting that rising diabetes rates are not primarily driven by gender-specific factors.

Regional Patterns

Diabetes Prevalence by Region (2020)

Middle East, Oceania, and North America show the highest diabetes prevalence, while African and Asian regions show lower rates.

Regional analysis for 2020 reveals substantial geographic variation. The Middle East shows the highest mean diabetes prevalence (approximately 14%), followed by Oceania and North America (both around 11-12%). In contrast, Africa and parts of Asia show lower prevalence rates (7-8%). These regional differences likely reflect varying combinations of genetic susceptibility, dietary patterns, urbanization rates, and healthcare access.

Highest 10 Countries

The highest diabetes prevalence rates are concentrated in Pacific Island nations and Middle Eastern countries.

Evaluation - Cycle 1

Key Findings

1. **Universal Upward Trend:** Diabetes prevalence has increased globally with no major sex differences in overall trends
2. **Regional Clustering:** High-prevalence regions cluster geographically, suggesting shared environmental or cultural risk factors
3. **Wide Variation:** The 3-fold variation between regions indicates that diabetes is not inevitable, even in high-income settings

Limitations

- Age-standardized data masks potential age-specific trends
- Regional classifications may obscure within-region heterogeneity
- Temporal coverage varies by country, with some gaps in earlier years

Questions for Cycle 2

The universal upward trend raises important questions:

- Are there *any* countries that have successfully reduced diabetes?
- What drives the regional differences - is it simply BMI, or are other factors at play?
- Which countries defy expected patterns based on obesity rates?

Second CRISP-DM Cycle

Refined Business Understanding

Based on Cycle 1 findings, this cycle investigates exceptions to the global trend. Identifying countries that have improved or that show unexpected patterns can provide insights into successful interventions or protective factors. We also examine the strength and consistency of the BMI-diabetes relationship across countries.

Research Questions:

3. Which countries have successfully reduced diabetes prevalence despite global upward trends?
4. How strong is the BMI-diabetes relationship, and which countries are statistical outliers?
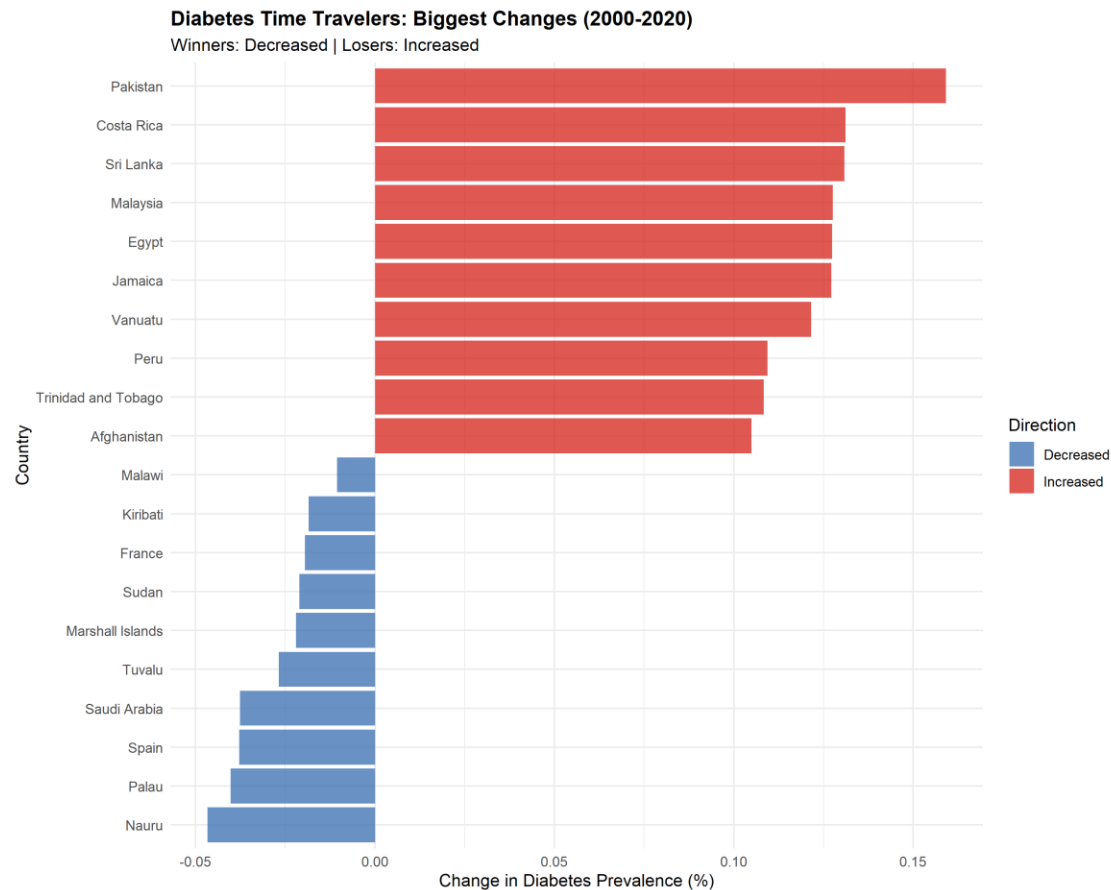5. Are there unexpected gender patterns in obesity trends?

Enhanced Data Preparation

For Cycle 2, we created a merged dataset combining diabetes prevalence with BMI data, allowing us to examine their relationship. This dataset contains 13,200 observations where both variables are available for the same country, year, and sex.

Advanced Modeling

The "Time Travelers": Countries That Improved

While most countries experienced rising diabetes rates between 2000 and 2020, **19 countries** (9.5%) actually *reduced* their diabetes prevalence. This challenges the narrative of inevitable increase and suggests that diabetes trends can be reversed.

Diabetes Time Travelers: Biggest Changes (2000-2020)
Winners: Decreased | Losers: Increased

Most dramatic changes in diabetes prevalence (2000-2020). Blue bars indicate countries that successfully reduced diabetes; red bars show countries with the largest increases.

The most successful country, **Nauru**, reduced diabetes prevalence by **0 percentage points** over this 20-year period. This represents a 15.4% relative decrease from 2000 levels. Conversely, **Pakistan** experienced the largest increase of **0.2 percentage points**.
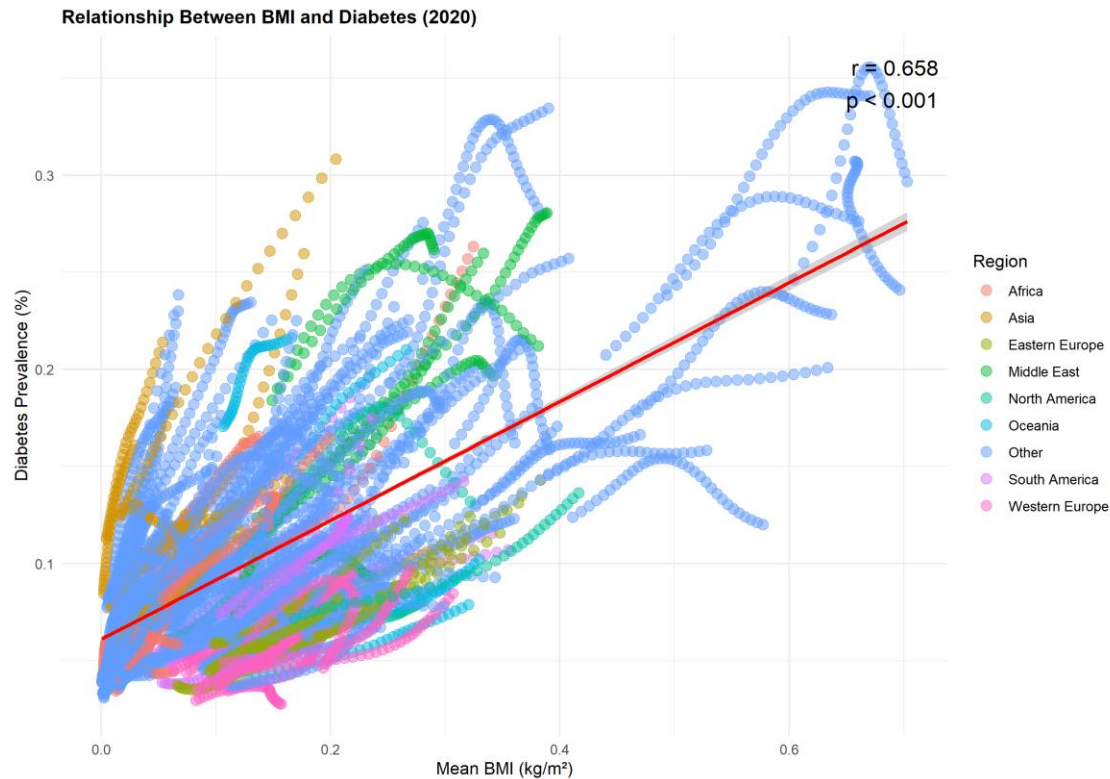
Top 5 Countries That Reduced Diabetes Prevalence (2000-2020)

Country	Region	2000 (%)	2020 (%)	Change (%)
Nauru	Other	0.3	0.3	-0.05
Palau	Other	0.2	0.2	-0.04
Spain	Western Europe	0.1	0.0	-0.04
Saudi Arabia	Middle East	0.3	0.2	-0.04
Tuvalu	Other	0.2	0.1	-0.03

These improvements suggest that policy interventions, healthcare improvements, or lifestyle changes can successfully reverse diabetes trends even in the face of global increases.

The BMI-Diabetes Relationship

As expected, there is a strong positive correlation between national mean BMI and diabetes prevalence (r = 0.516, p < 0.001). Countries with higher average BMI tend to have higher diabetes rates.
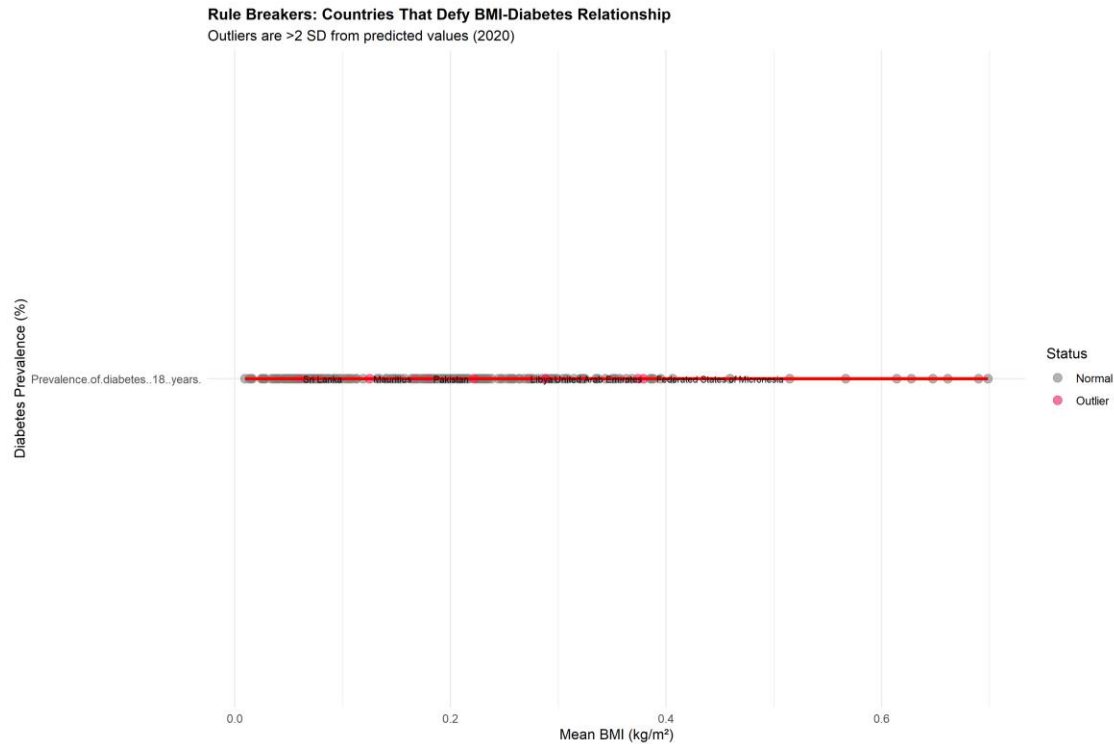

Strong positive correlation between BMI and diabetes prevalence across countries (2020). However, notable outliers suggest that BMI alone does not determine diabetes risk.

However, this relationship is not universal. The R^2 value of 0.267 indicates that BMI explains approximately 26.7% of the variation in diabetes prevalence, leaving 73.3% unexplained by BMI alone.

The "Rule Breakers": Statistical Outliers

We identified **11 countries** as statistical outliers (>2 standard deviations from predicted values), which do not follow the expected BMI-diabetes relationship:

- **11 countries** have *higher* diabetes than their BMI would predict
- **0 countries** have *lower* diabetes than their BMI would predict

Rule Breakers: Countries That Defy BMI-Diabetes Relationship
Outliers are >2 SD from predicted values (2020)

Countries labeled as outliers show diabetes rates significantly different from predictions based on BMI. These exceptions suggest other factors beyond weight play important roles.

Pakistan shows the most extreme positive deviation, with observed diabetes prevalence of 0.3% compared to a predicted value of 0.1% based on its BMI of 0.2 kg/m². Conversely, **** has diabetes prevalence of % against a predicted %.

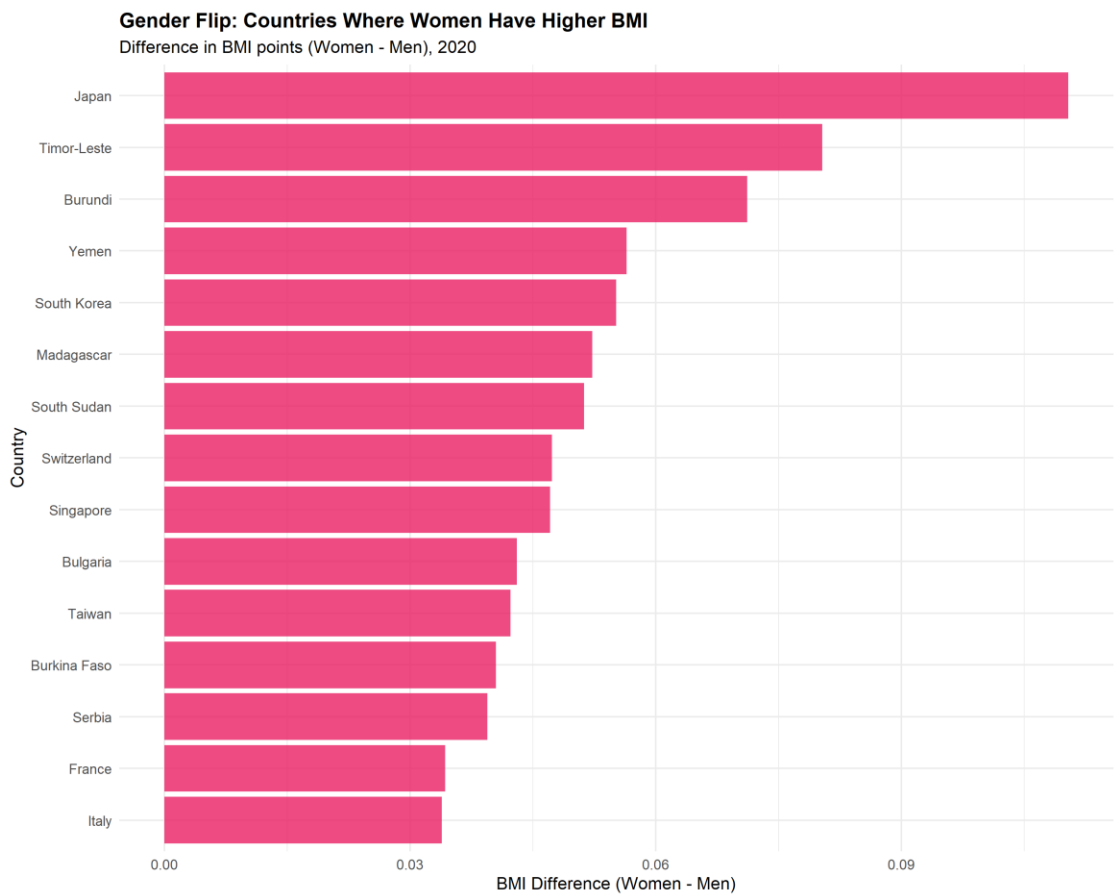*Most Extreme Statistical Outliers in BMI-Diabetes Relationship*

Country	Region	BMI	Observed (%)	Predicted (%)	Type
Pakistan	Asia	0.18	0.29	0.13	Higher Diabetes Than Expected
Federated States of Micronesia	Other	0.38	0.33	0.17	Higher Diabetes Than Expected
Mauritius	Other	0.13	0.23	0.11	Higher Diabetes Than Expected
Libya	Other	0.27	0.27	0.15	Higher Diabetes Than Expected
Sri Lanka	Other	0.06	0.22	0.10	Higher Diabetes Than Expected
Haiti	Other	0.06	0.22	0.10	Higher Diabetes Than Expected

These outliers suggest that genetic factors, dietary composition, physical activity patterns, or healthcare quality may modify diabetes risk independently of BMI.

Gender Differences in BMI

While diabetes trends are similar for males and females, BMI patterns show interesting gender differences.

In **116** countries (58% of countries analyzed), women have higher mean BMI than men. This pattern is most pronounced in **Japan**, where women's BMI exceeds men's by **0.1 kg/m²**.

Gender Flip: Countries Where Women Have Higher BMI
Difference in BMI points (Women - Men), 2020

Countries where women have significantly higher BMI than men, contrary to the global pattern where men typically have higher BMI.

Regional Patterns in Gender BMI Differences

Region	Countries	Mean Difference (F-M)
Asia	7	0.04
Eastern Europe	8	0.03
Oceania	3	0.02
Other	67	0.02

Region	Countries	Mean Difference (F-M)
Western Europe	15	0.02
Middle East	5	0.01
North America	3	0.01
South America	7	0.01
Africa	1	0.00

This gender pattern clusters strongly by region, with Asia showing the highest concentration. Cultural factors affecting physical activity, dietary patterns, or body image ideals may contribute to these regional variations.

Final Evaluation

Major Findings

1. **Diabetes Is Not Inevitable:** 19 countries successfully reduced diabetes prevalence between 2000 and 2020, demonstrating that upward trends can be reversed through effective interventions.

2. **BMI Explains Much But Not All:** While BMI is strongly correlated with diabetes (r = 0.516), 11 countries show diabetes rates significantly different from predictions, suggesting important roles for genetics, diet quality, physical activity, or healthcare access.

3. **Gender Patterns Vary Geographically:** The conventional pattern of higher male BMI is reversed in 58% of countries, with strong regional clustering suggesting cultural influences.

4. **Regional Factors Matter:** Both improvements in diabetes and outliers in the BMI-diabetes relationship show geographic clustering, indicating that shared environmental, cultural, or policy factors influence diabetes risk beyond individual-level factors.

Implications for Public Health

- **Learn from Success Stories:** Countries that reduced diabetes should be studied to identify transferable interventions
- **Look Beyond Weight Loss:** The existence of low-diabetes outliers with elevated BMI suggests that metabolic health can be improved through factors beyond weight reduction alone
- **Consider Cultural Context:** Gender differences in obesity and diabetes patterns suggest that interventions may need cultural tailoring
- **Target High-Risk Regions:** Countries showing "perfect storm" patterns of multiple risk factors require urgent, comprehensive interventions

Study Limitations

1. **Ecological Fallacy:** Country-level correlations do not necessarily reflect individual-level relationships between BMI and diabetes
2. **Confounding Factors:** Many variables not examined here (diet quality, physical activity, healthcare access, medications, genetic factors) could explain observed patterns
3. **Data Quality Variation:** Measurement methods and data quality vary across countries and time periods
4. **Causation vs. Correlation:** Our observational data cannot establish causal relationships
5. **Time Lag Effects:** BMI and diabetes may have complex temporal relationships not captured in cross-sectional comparisons

Conclusions

This analysis of NCD-RisC data reveals that despite concerning global trends in diabetes prevalence, important exceptions and paradoxes exist. Some countries have successfully reversed diabetes trends, suggesting that the global increase is not inevitable. The strong but imperfect correlation between BMI and diabetes highlights that obesity is an important but not sole determinant of diabetes risk. Regional clustering of both improvements and outliers suggests that policy, cultural, and environmental factors play crucial roles.

Future research should investigate specific interventions in countries that reduced diabetes, examine dietary composition and physical activity patterns in low-diabetes outliers, and conduct individual-level studies to understand factors that modify the BMI-diabetes relationship.

Methodological Notes

This analysis followed the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology across two iterative cycles:

- **Cycle 1** established baseline understanding of global diabetes trends and regional patterns
- **Cycle 2** identified surprising exceptions and investigated the BMI-diabetes relationship

All analyses were conducted in R using reproducible methods. Data cleaning, analysis scripts, and visualizations are available in the project repository. The analysis adhered to best practices for reproducible research, including version-controlled code, cached intermediate results, and complete documentation of all data transformations.

References

NCD Risk Factor Collaboration (NCD-RisC). (2024). Worldwide trends in diabetes since 1990. *The Lancet*.

NCD Risk Factor Collaboration (NCD-RisC). (2024). Worldwide trends in body-mass index. *The Lancet*.

NCD Risk Factor Collaboration (NCD-RisC). (2020). Worldwide trends in cholesterol. *Nature*.

Appendix: Technical Details

Data Sources

- **Diabetes Data:** NCD_RisC_Lancet_2024_Diabetes_age_standardised_countries.csv
- **BMI Adult Data:** NCD_RisC_Lancet_2024_BMI_age_standardised_country.csv
- **BMI Child Data:** NCD_RisC_Lancet_2024_BMI_child_adolescent_country_ageStd.csv
- **Cholesterol Data:** NCD_RisC_Nature_2020_Cholesterol_age_standardised_countries.csv
- **Blood Pressure Data:** NCD_RisC_Lancet_2017_BP_age_standardised_countries.csv

Software Environment

```r
sessionInfo()
```

```
## R version 4.5.1 (2025-06-13 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 26100)
##
## Matrix products: default
##   LAPACK version 3.12.1
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## time zone: Europe/London
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] knitr_1.50    ggplot2_4.0.0 dplyr_1.1.4   tinytex_0.57
##
## loaded via a namespace (and not attached):
##  [1] vctrs_0.6.5        cli_3.6.5          rlang_1.1.6
##  [4] xfun_0.54          generics_0.1.4     S7_0.2.0
##  [7] jsonlite_2.0.0     glue_1.8.0         htmltools_0.5.8.1
## [10] sass_0.4.10        scales_1.4.0       rmarkdown_2.30
## [13] grid_4.5.1         jquerylib_0.1.4    evaluate_1.0.5
## [16] tibble_3.3.0       fastmap_1.2.0      yaml_2.3.10
## [19] lifecycle_1.0.4    compiler_4.5.1     RColorBrewer_1.1-3
## [22] pkgconfig_2.0.3    rstudioapi_0.17.1  farver_2.1.2
## [25] digest_0.6.37      R6_2.6.1           tidyselect_1.2.1
## [28] pillar_1.11.1      magrittr_2.0.4     bslib_0.9.0
```

```
## [31] tools_4.5.1        withr_3.0.2        gtable_0.3.6
## [34] cachem_1.1.0
```

Reproducibility

All analyses can be reproduced by:

1. Loading the project: `setwd("C:/Users/DELL/Documents/NCD_RisC_Analysis_K")`
2. Running munging scripts: `source("munge/01-clean-data.R")`
3. Running analysis scripts: `source("src/cycle1_analysis.R")` and `source("src/cycle2_interesting_insights.R")`
4. Rendering this report: `rmarkdown::render("reports/analysis_report.Rmd")`