Study Title: Epidemiologic study to Investigate the Prevalence of Insulin Resistance and associated Cardiovascular Diseases in the Metropolitan area of Naples (Italy):

Date: January 20,2026

NCT number: pending

## Study protocol and statistics

The study originated from a collaboration among the TDC19 foundation, the Municipality of Naples, the ASL Napoli 1 Centro (Local Health Authority), the University of Naples Federico II, and the COMEGEN general practitioner cooperative.

## Materials and methods

### *Study design, setting, and data source*

This was a retrospective cross-sectional observational study based on routinely collected primary-care data. The data source was the COMEGEN Cooperative database, which includes general practitioners affiliated with ASL Napoli 1 Centro and contains clinical and laboratory information for approximately 200,000 individuals distributed across municipal districts of Naples.

For the present analysis, the most recent available records within the six months preceding data extraction were screened. Participants were eligible if they were aged 18-60 years, had no recorded overt diabetes, and had fasting glucose and fasting triglyceride values obtained from the same assessment window, allowing TyG calculation.

### *Participants and sampling*

The analytic cohort was a convenience cohort including all eligible adults with complete data for TyG calculation. Because laboratory availability was required, the cohort should be interpreted as a large primary-care sample rather than a probability sample of all residents. Selection mechanisms related to clinical testing, health-care access, socioeconomic status, and comorbidity were considered potential sources of selection bias.

### *Variables and definitions*

Demographic, anthropometric, biochemical, clinical, and geographic variables were extracted from the COMEGEN primary-care database. Demographic variables included age, defined as age at the index laboratory assessment, and sex, as recorded in the general-practitioner database. Anthropometric data included body mass index (BMI), calculated as weight in kilograms divided by height in meters squared. Biochemical variables included

fasting plasma glucose, triglycerides, total cholesterol, high-density lipoprotein cholesterol, and low-density lipoprotein cholesterol, all expressed in mg/dL. The triglyceride-glucose index was calculated using fasting triglyceride and fasting glucose values according to the prespecified formula: TyG = ln[(fasting triglycerides (mg/dL) × fasting glucose (mg/dL))/2]. TyG-defined insulin resistance was defined as a TyG value ≥8.5. (17-19). Clinical information included the presence of recorded cardiovascular disease, as documented in the primary-care records. Geographic information included the municipal district of residence within the municipality of Naples; availability and missingness of district-level information were recorded for the district-level analysis.

**Statistical analysis**

Continuous variables were summarized as mean +/- standard deviation (SD) or median and interquartile range (IQR), depending on distributional characteristics. Distributions were assessed by graphical inspection and, where appropriate, formal normality testing; given the large sample size, clinical relevance and distributional shape were prioritized over normality-test p-values alone.

Age-group comparisons were performed using analysis of variance for continuous variables and Pearson chi-square tests for categorical variables. The prevalence of TyG-defined IR was reported with a 95% confidence interval (CI).

Associations with TyG-defined IR were evaluated using logistic regression. Univariable models were fitted, followed by a multivariable model adjusted for age, sex, BMI, and recorded CVD. Results are reported as odds ratios (ORs) with 95% CIs. Because recorded CVD was modeled as a predictor of TyG-defined IR, the corresponding OR should be interpreted as the cross-sectional association between CVD and IR, not as the effect of IR on incident CVD. A complementary model with recorded CVD as the dependent variable and TyG-defined IR as the main exposure would be required to address the manuscript title more directly.

Missing data were handled by complete-case analysis. All tests were two-sided and $p<0.05$ was considered statistically significant. Analyses were performed using Stata version 18.

***Sample size and ethics***

The minimum required sample size for estimating an expected IR prevalence of 30% with 1% absolute precision at a 95% confidence level was 8,068 participants. The final analytic

sample of 17,510 participants was therefore adequate for prevalence estimation, although representativeness depends on the completeness and selection mechanisms of available laboratory data.

Physicians obtained informed consent from patients for the anonymized use of their data. The study was approved by the Campania I Territorial Ethics Committee on January 20, 2026, and was conducted in accordance with the principles of the Declaration of Helsinki.