Implementation of the Super Protected protocol to increase HPV vaccination in girls and adolescents aged 9 to 14 years in the municipality of Envigado: A Behavioral Randomized Controlled Trial

Statistical analysis plan

Database creation. The data was collected through a web form in the Google Forms application, so that a database was created in the Microsoft Excel program. In order to control possible errors or omissions, the process of making each question obligatory to answer in order to move on to the next one was activated and, in turn, the option was activated that made it possible, for example, to restrict the type of data that could be entered in each variable (only numbers, letters or dates).

Database debugging. Once the database was set up, it was examined to check for possible errors or duplicate records and, once it was cleaned of inconsistencies, it was exported to the statistical package SPSS (Statistical Package for the Social Sciences) version 24.0 for analysis.

Treatment of missing data. If incomplete records were found, the participant was contacted by telephone to request the missing data; if it was not possible to contact them, the percentage of missing information was analyzed and if it was higher than 20%, the record was eliminated.

Statistical analysis. It comprised the following parts: Descriptive analysis: a descriptive analysis of the characteristics of the sample, according to its nature and level of measurement, in the case of qualitative variables, frequencies were described in terms of percentages and quantitative variables in measures of central tendency (mean, median and interquartile range).

Pre-post bivariate analysis. The analysis of association between cases and controls was performed with Fisher's Chi-square test, in the case of categorical variables, and Student's t-tests or Man Whitney U-tests, in the case of continuous quantitative variables.

Intragroup repeated-measures analysis was also performed with McNeimar's Chi-square test for categorical variables and Student's t-test for repeated samples or Wilcoxon's test for continuous variables.

Phase 1: Baseline measurement, at the beginning of the study: the groups were analyzed according to the nature and level of measurement of each variable. To contrast variables of a qualitative nature.

Phase 2: Multivariate analysis (intervention effect measure), for this phase a multivariate logistic regression analysis was performed to calculate the probability (OR, 95% CI) of HPV vaccination in both the experimental and control groups, controlling or taking into account those intermediate variables that were associated in the bivariate analysis (e.g., sociodemographic data).

Finally, measures of association and impact of the intervention were calculated by calculating the relative risk (RR), relative risk reduction (RRR), absolute risk reduction (ARR) and number needed to treat (NNT); all these estimators with the 95% CI. A significance level of $p<0.05$ was assumed for all contrasts.