

**Seroprevalence of Pertussis among Healthy Children and
Adolescents in Kazakhstan, January-October 2021**

STATISICAL ANALYSIS PLAN

Official Title of the study:

**«Seroprevalence of Pertussis among Healthy Children and
Adolescents in Kazakhstan: A Cross-Sectional Study»**

Date of the document: **October 09, 2020**

NCT number: n/a

Internal Sanofi Pasteur Study Code: **PER00075**

Scientific organization: **Scientific and practical centre of sanitary and
epidemiological expertise and monitoring, Almaty, Kazakhstan**

Seroprevalence of Pertussis among Healthy Children and Adolescents in Kazakhstan, January-October 2021

Statistical analysis plan (PER00075, Point 7)

This is a cross-sectional sero-survey study in a population sample aged ≥ 10 and < 19 years of age. The data will be analysed descriptively..

This is a cross-sectional sero-survey study in a population sample aged ≥ 10 and < 19 years of age. The data will be analyzed descriptively. The database will contain several sections with different variables: for demographic, socio-economic, vaccination status, and infection risk factors. The questionnaires will be grouped by grade and section to facilitate data entry. Information from the questionnaires will be entered into an online form created using Google Forms by specialists in 4 regions. Input inconsistencies found will be corrected by rechecking the relevant questionnaires until a clean database will be received. The data will be validated according to a logical sequence of questions. For this purpose, prime frequencies will be calculated for all variables, allowing inconsistencies and incomplete information to be detected. Any discrepancies found will be discussed by the principal investigators to determine the action to be taken, and changes will be made by a single data scientist. This completes the data cleansing process.

EpiInfo 7 software for Windows will be used to calculate simple averages and frequencies, taking into account a confidence interval (CI) of 95% and to determine the statistical significance of the results obtained ($p < 0.05$). Descriptive statistical tests, proportions and mean will be used to calculate demographic and socioeconomic factors, immune status and risk of infection.

The analysis will be conducted using the Statistical Package for the Social Sciences (SPSS). Correlations between age group and anti-RT antibody levels will be analyzed using multivariate regression or regression splines. Binary and multivariate logistic regression analysis will be used to determine the relationship and influence of independent variables on the dependent variable.