Immunity in the Swedish population: diphtheria, tetanus and poliomyelitis

Åke Svensson, a Margareta Böttiger b and Olle Gustavsson c

Background
Large-scale vaccination programmes have been carried out for a long time in Sweden, as in many other countries. However, often little is known of the effects of these vaccinations. During 1990 and 1991 a survey of immunity based on a random adult population sample was carried out. The main purpose was to estimate the level of immunity to diphtheria, tetanus and polio of the adult population in Sweden. In total 4800 people were randomly selected according to a stratified, two-stage, sampling plan.

Methods
Based on standard sampling theory, methods for calculations of estimates and confidence intervals of the proportion of the population that is immune are given. The response patterns and its possible effects on the estimates are discussed.

Results
In total, 70.6% of the 4800 selected gave a blood sample. The response rate differs for men and women and for different age groups. Among the oldest, the response rate was close to 80%, but it was only about 60% among the youngest.

Conclusions
With the survey design used, it was possible to obtain a sufficient degree of response. Our experience is that the response rate depends to a large extent on the efforts made to explain and motivate participation.

Keywords
Vaccination, immunity, serology

Accepted
2 February 1998

Large-scale vaccination programmes have been carried out for a long time in Sweden, as in many other countries. However, often little is known of the effects of these vaccinations. It is not an easy task to derive information about the actual level of immunity of the population to infections of different kinds. One way is to keep careful registers of the vaccinations made. Together with a follow-up of the quality and effectiveness of the vaccines used, such data can provide a general picture of the situation. However, such registers do not exist, except perhaps for children vaccinated in the child-care service or in school. Knowledge of the effects of the vaccinations and the duration of immunity after the different vaccination schedules that have been used over the years is scanty. Another method sometimes used is based on antibody testing of blood samples that, for one reason or another, have been obtained and stored in the health-care system. This is a simple and cheap way of obtaining blood samples. The drawback is that there is no guarantee of representativeness and there are no possibilities of judging the precision or certainty of the results obtained. There are also a few studies, based on samples from special groups of individuals, from which blood samples are easy to access. Examples of groups that have been used in different studies are pregnant women and military personnel. Such samples suffer from the same defects, as regards representability and quality.

A straightforward procedure for getting a trustworthy picture of the immunity in the entire population is to analyse sera from a random sample of the population. This is, of course, in many ways a difficult method, and to our knowledge there are few large-scale surveys of this kind. In 1968, a survey with the sole purpose of studying immunity based on a random sample of the Swedish population was carried out.1 Its aim was to evaluate the general Swedish programme of vaccination against polio, a programme that started in 1957. Another possibility is to analyse blood samples obtained in surveys which have been made with a broader purpose. The immunity status in the population can then be obtained as an (important) by-product. As part of the third National Health and Nutrition Examination Survey in the USA, immunity to tetanus was estimated from blood samples.2 In this study, 77.4% of people interviewed for the survey consented to give a blood sample.

To carry out a survey of the population based on a random sample requires an organization able to collect and analyse blood samples and to report and interpret the results. There is always a risk of a large percentage of non-participants. People selected for the sample have to be asked to contribute a blood sample for reasons that are not directly connected with their
The 1968 survey of immunity to polio\(^1\) was designed as a stratified sample divided into 16 age groups with two parishes selected in each of the three largest cities, 10 of the remaining cities and 10 rural communities. At that time, some regions were excluded from the survey, as it was supposed that their inhabitants would not be willing to participate in the survey. The sample consisted of 3031 people, of whom 90% answered a questionnaire and 75.7% participated in the blood-sampling. This survey yielded interesting results and had an important influence on the continuation of the Swedish programme of vaccination against polio.

During 1990 and 1991, the National Bacteriological Laboratory in Sweden (SBL) (since then re-organized as the Swedish Institute for Infectious Disease Control) planned and carried through a new survey of immunity based on a random sample. The main purpose was to estimate the level of immunity to diphtheria, tetanus and polio of the adult population in Sweden. In this paper, the planning of the survey, the participation and the methods by which the results were analysed are described. The results and conclusions concerning the particular infections are presented in a separate article.\(^3\)

The participants agreed that the sera could be used, after making identification impossible, to search for antibodies against other infections than those of primary interest, with the explicit exception of HIV. The survey was considered and confirmed by the Ethical Committee of the Karolinska Institute and by the Swedish Data Inspection Board (Datainspektionen).

## Materials and Methods

### Population studied

The reference population was people born before 1973 and entered in a parish register. In Sweden, all permanent residents are entered in a parish register, regardless of citizenship. The survey was restricted to the adult population, since it was thought that data on immunity and vaccination for individuals under 18 years of age could be more easily be derived from other sources of information, such as the child and school health service. The total number of people eligible for the survey was about 6 million.

### Survey design

Experience from the previous study of immunity to polio\(^1\) and other surveys of this kind shows that considerable work has to be devoted to finding the selected people and motivating them to contribute a blood sample. In order to make this possible without building up a separate and expensive organization, each person involved locally should only be responsible for obtaining a limited number of blood samples. Sweden is divided into 24 counties. In each of these counties, a county medical officer (smittskyddsläkare) is responsible for the local supervision of communicable diseases. It was decided that the sample should roughly be stratified over the 24 counties in proportion to the size of the population and that the collection of blood samples should be made under the supervision of the responsible county medical officers.

The survey was planned to give results with regard to the proportion of the population that had antibodies exceeding some titre with a precision of about ±5%, when divided according to sex and year of birth (five age groups) or divided according to sex in a geographical subdivision of Sweden into four different parts. The four geographical areas are referred to as Southern, Western, Central and Northern Sweden (Table 1).

In total, 4800 people, randomly selected according to a stratified, two-stage, sampling plan, were contacted and asked to contribute a blood sample. Those eligible were selected from the Swedish register of people and addresses (SPAR). The basic facts about the counties as regards total population, number of parishes, number of selected parishes, number of people in the sample and number of responses are given in Table 1.

In the first stage of the sample, a number of parishes were selected in each county. These parishes were selected in such a way that their chance of being included in the sample was roughly proportional to their size (i.e. the total number of inhabitants in the parish). In the selection procedure actually used, the parishes were selected one by one with a probability proportional to their size and without replacement until the prespecified number of parishes was obtained in each county. One of the counties (Gotland) is extremely small and, according to the design, only one parish should be selected. Gotland consists of 92 parishes; the largest parish (Visby) has about 40% of the total population and the others are all fairly small. It was decided that Visby should be included and one of the other parishes and that the number of selected people in the two selected parishes should be half (24 people), compared with the others.

In the second stage, a sample was obtained by systematic sampling from a list of the people registered in the parish in order of date of birth. The sample is stratified in such way that four people of each sex born before 1930, between 1931 and 1945 and between 1946 and 1955 and six people born between 1956 and 1965 and between 1966 and 1972 were selected.

### Organization

Those selected to be included in the study were informed by letters from their county medical officer. Information about the study was also spread via the media, local newspapers, radio and television. The general plan did not include the collection of any other information from the participants besides the blood sample and the data on date of birth and present address provided by the register used for the random selection. The only personal benefit for the participants was that they were told about their immunity status in relation to diphtheria, tetanus and polio types 1, 2, and 3. Those lacking immunity were offered vaccination.

### Handling and analysis of the blood samples

The blood samples were collected county by county by the local primary healthcare organizations and sent consecutively overnight by regular mail to the National Bacteriological Laboratory (SBL). At the Laboratory, the sera samples were frozen to −70°C until they were analysed. The sera were thawed and analysed for antibodies against diphtheria and tetanus toxins and the three types of polio virus.

For detailed accounts of the serological assays and the choice of levels of antibodies that were assumed to afford protection, the reader is referred to the special report.\(^3\)
The principal aim of the study was to derive estimates of the proportion of the inhabitants, divided according to year of birth, sex and place of residence, that have antibodies above some threshold level against the infections under consideration. For simplicity we shall refer to those with sufficient antibody levels as immune. The precision of the estimates was to be described according to the group in the parishes and the inclusion probability of the immune among the participants. These estimates are combined to form an estimate for the entire country or for one of the geographical areas by weighting these crude estimates with weights depending on the actual number of inhabitants belonging to the group in the parishes and the inclusion probability of the parish. For details of how this weighting is done, the reader is referred to the Appendix.

Non-participation is treated as uninformative, i.e. we have postulated that non-response is due to chance and is not related to the presence or absence of antibodies in the blood. In order to illustrate how sensitive the results are to this assumption, estimates are also calculated on a crude assumption of how the willingness to participate in the survey differs as between the immune and non-immune people.

### Results

#### Response and non-response

In total, 3390 (70.6%) of the 4800 selected gave a blood sample. The response rate differs for men and women and for the different age groups (Table 2). Among the oldest, the response...
rate was close to 80%, but was only about 60% among the youngest.

There are different reasons for not participating in the investigation. A smaller fraction of the non-response is explained by deficiencies in the sample frame. Some of the selected people have moved to other places or are deceased. However, many people will not, for one reason or another, consent to giving a blood sample. A closer examination showed that the willingness to respond varied between counties and parishes. The variations may, to some extent, be due to geographical and sociological differences between the counties, but it is our impression that the response rate obtained depended substantially on the efforts made by the local organizers and collectors. Another cause of variation between the number of responders in different parishes is the sampling procedure.

Let us suppose that the proportion, $p$, of the inhabitants of a certain parish are potential responders. Then there will be a variation in the actual number of responders due to the random selection of individuals in the sample. This variation can be described to a sufficient degree of approximation by a binomial distribution with $p$ as the probability parameter. By comparing the variation in response rate in different parishes with the theoretical variation that would follow from a binomial variation (with common $p$), we can describe the variation in willingness to respond between parishes. This variation can be measured by comparing the actual variance of response rates with the variance predicted by a binomial model. The ratio of the two variances is called over-dispersion. In Table 3, the distribution of the number of responses is given and the over-dispersion is calculated. The over-dispersion varies between the groups. It is clear that there is a larger extra variation in response rates for males than for females for each age group and that the variation in response rate is larger for younger cohorts. This indicates that males and young people are more sensitive to extra efforts made to promote participation.

There is also a clear correlation of the response rates for different age groups, i.e. parishes tend to have either low or high response rates in all age groups. In fact, variation in response rates between parishes explains 74% of the variation in the total number of responses in the parishes.

**Response and immunity**

If the willingness to participate in the survey is correlated with the immunity status of people, this will cause considerable difficulties in analyzing and interpreting the results. There are some reasons why this could be the case. People unwilling to be vaccinated could also be unwilling to give a blood sample. If this is the case, the immune are over-represented in the survey. However, there may also be causes working in the opposite direction. A person suspecting that his or her protection is not satisfactory may be willing to take the opportunity offered to investigate this. In short, there may be confounding factors influencing both the willingness to respond and the immunity.

In short, the immune may be over-represented in the survey. However, there may also be causes working in the opposite direction. A person suspecting that his or her protection is not satisfactory may be willing to take the opportunity offered to investigate this. In short, there may be confounding factors influencing both the willingness to respond and the immunity.

There is no entirely convincing way of concluding whether if this is the case from the study of the sample itself.

The most pessimistic assumption would be that all non-responders, in fact, do not have a sufficient level of antibodies. A more realistic model can be based on the assumption that

---

**Table 2** Response rates by region, year of birth and sex (in %)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Men</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>

**Table 3** Distribution of number of responses in parishes (the two parishes in county 9 are aggregated)

<table>
<thead>
<tr>
<th>No. of responses</th>
<th>Women</th>
<th>Men</th>
<th>Women</th>
<th>Men</th>
<th>Women</th>
<th>Men</th>
<th>Women</th>
<th>Men</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>12</td>
<td>1</td>
<td>10</td>
<td>4</td>
<td>14</td>
<td>3</td>
<td>8</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>16</td>
<td>7</td>
<td>12</td>
<td>14</td>
<td>23</td>
<td>10</td>
<td>19</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>23</td>
<td>12</td>
<td>19</td>
<td>38</td>
<td>24</td>
<td>40</td>
<td>40</td>
<td>36</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>18</td>
<td>25</td>
<td>27</td>
<td>42</td>
<td>38</td>
<td>47</td>
<td>33</td>
<td>41</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>16</td>
<td>27</td>
<td>21</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>10</td>
<td>28</td>
<td>10</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Mean 3.84 3.25 4.54 3.64 3.16 2.84 3.31 2.98 3.13 3.21
Variance 0.64 0.54 0.76 0.61 0.79 0.71 0.83 0.75 0.78 0.80
Over-dispersion 1.38 1.84 1.44 1.54 1.20 1.50 1.04 1.10 1.14 1.27
the willingness to participate differs between immune and non-immune people. Let us assume that \( p \) is the probability that an immune person in a certain parish is willing to give a blood sample and that \( \bar{p} \) is the corresponding probability for a non-immune person in the same parish. Let us assume that there is a constant ratio between these probabilities, i.e. \( p = \delta \bar{p} \). Suppose that 80% of the immune in a parish are willing to participate in the study, then, if \( \delta = 0.8 \), 64% of the non-immune will respond. If \( \delta = 0.5 \), only 40% of the non-immune will, in this case, give a blood sample. Let \( I \) denote the number of immune responders and let \( A \) denote the total number of responders in the parish. If the non-response is uninformative, i.e. \( \delta = 1 \), then \( I/A \) is an unbiased estimate of the proportion of immune in the parish. If \( \delta < 1 \), \( I/A \) will be an overestimate. The size of the bias depends on the proportion of immune. However, simple calculations yield that the bias will be considerable less. The sensitivity of the estimates to different values of \( \delta \) is illustrated in Figures 1 and 2. In Figure 1, the proportions of females and males immune to polio type 3 are estimated for different \( \delta \)-values. In Figure 2, the same is done for the proportions of females and males born between 1946 and 1955 immune to diphtheria. The two situations illustrated differ in the proportion of the population that has antibodies against the infection. The diphtheria example illustrated in Figure 2 is the case that turned out to be most sensitive to response bias. The bias for reasonable \( \delta \) is of the size of the standard deviation of the estimate. Even though it should be kept in mind in interpreting the data, the deviation can be considered as small for most purposes.

Crude estimates

The simplest way to estimate immunity from the result of the survey would be to calculate crude ratios, i.e. to calculate the number of immune among the responders. The survey was planned in such a way that such self-weighted estimates would not be seriously biased. The differences between crude (un-weighted) estimates and the weighted estimates are illustrated in Table 4. It will be seen that in all cases the differences are small. However, it should be observed that the distribution over the five cohorts according to year of birth does not reflect the relative proportions of people in these cohorts. Crude estimates that are not stratified according to age may be seriously biased.

Table 4 Weighted and crude estimates of proportion of immune (in %) according to year of birth

<table>
<thead>
<tr>
<th>Born during</th>
<th>Tetanus</th>
<th>Diphtheria</th>
<th>Polio type 1</th>
<th>Polio type 2</th>
<th>Polio type 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Weighted</td>
<td>Crude</td>
<td>Weighted</td>
<td>Crude</td>
<td>Weighted</td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1966-1972</td>
<td>98.6</td>
<td>98.7</td>
<td>93.9</td>
<td>94.2</td>
<td>98.8</td>
</tr>
<tr>
<td>1956-1965</td>
<td>88.3</td>
<td>87.9</td>
<td>77.6</td>
<td>76.0</td>
<td>98.8</td>
</tr>
<tr>
<td>1946-1955</td>
<td>53.5</td>
<td>54.4</td>
<td>40.4</td>
<td>39.6</td>
<td>97.6</td>
</tr>
<tr>
<td>1931-1945</td>
<td>44.3</td>
<td>43.9</td>
<td>45.4</td>
<td>45.9</td>
<td>99.1</td>
</tr>
<tr>
<td>&lt;1930</td>
<td>30.3</td>
<td>30.4</td>
<td>25.8</td>
<td>24.6</td>
<td>99.8</td>
</tr>
<tr>
<td>Men</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1966-1972</td>
<td>98.6</td>
<td>98.8</td>
<td>93.4</td>
<td>94.2</td>
<td>98.8</td>
</tr>
<tr>
<td>1956-1965</td>
<td>95.4</td>
<td>95.6</td>
<td>87.3</td>
<td>86.8</td>
<td>95.8</td>
</tr>
<tr>
<td>1946-1955</td>
<td>89.9</td>
<td>90.1</td>
<td>61.1</td>
<td>60.2</td>
<td>95.4</td>
</tr>
<tr>
<td>1931-1945</td>
<td>83.2</td>
<td>82.6</td>
<td>64.8</td>
<td>63.8</td>
<td>95.9</td>
</tr>
<tr>
<td>&lt;1930</td>
<td>50.1</td>
<td>50.5</td>
<td>24.6</td>
<td>24.6</td>
<td>93.0</td>
</tr>
</tbody>
</table>
Discussion
The planning of the survey differed in several aspects from the previous polio investigation. Though great care was taken in planning and advertising the present survey, it was not at all clear that it would be possible to successfully carry through a survey of this kind. Many things have changed since 1968 as regards the general willingness to participate in random surveys. The HIV/AIDS epidemic may also make individuals hesitate to give blood samples for the detection of antibodies.

The study has shown that it is in fact possible to get qualitatively good estimates of immunity by using a random sample. With the survey design used, it was possible to obtain a sufficient degree of response. Our experience is that the response rate depends to a large extent on the efforts made to explain and motivate participation. It can be noted that it was possible in some sparsely populated areas to obtain almost total participation of those selected. One main conclusion is that it was possible—in spite of many obstacles—to carry out this study on a strictly statistical basis.

In comparison with the response rate of other surveys, the degree of participation was viewed as satisfactory. The analysis of the results and the conclusions that can be drawn from the survey will, of course, suffer from the fact that the response was not total.

The results for the particular infections analyses are presented in a separate article.\(^1\) As could be expected, considerable differences in serological immunity according to both age and place of residence are observed. Thus the survey has contributed valuable information on the population’s immunity status.

The blood samples are now stored, after removing any information that would make personal identification possible. They can in the future be used to analyse immunity to other infections. Such analyses have to be carried out with some of the information connected with the individual blood samples missing. The self-weighting property will make it possible to give reasonable estimates even in this case.

This material gives a cross-sectional picture of the situation in the Swedish population in 1990 and 1991. It is worthwhile considering the possibilities of repeating studies of this kind in the future, in order to investigate changes in the situation.

References
5 Raj D. Sampling Theory. New York: McGraw-Hill 1968, p.120.

Appendix
Calculation of estimates and confidence intervals
We shall use the following notation (\(r\) stands for region, \(c\) for county, \(p\) for parish, \(g\) for age- and sex-group, and \(G\) for an aggregation of groups):

\[
\begin{align*}
N_c & : \text{Number of inhabitants in county } c, \\
N_{cp} & : \text{Number of inhabitants in parish } p \text{ in county } c, \\
N_{cp,g} & : \text{Number of inhabitants in parish } p, \text{ county } c \text{ in group } g, \\
\theta_{cp,g} & : \text{Proportion of immune inhabitants in parish } p, \text{ county } c \text{ in group } g, \\
D_c & : \text{Number of parishes in county } c \text{ sampled}, \\
A_{cp,g} & : \text{Number of responders in county } c, \text{ parish } p, \text{ and group } g, \\
I_{cp,g} & : \text{Number of immune responders in county } c, \text{ parish } p \text{ and group } g.
\end{align*}
\]

The proportion of immune individuals in region \(r\) belonging to group \(G\) can, using these notations, be expressed as:

\[
\begin{align*}
\hat{\theta}_{rg} = & \frac{\sum r \sum c \sum p \sum g N_{cp,g} \theta_{cp,g}}{\sum r \sum c \sum p \sum g N_{cp,g}}. \\
\end{align*}
\]

The estimate of the proportion of immune in region \(r\) belonging to group \(G\) is built up of a ratio between an estimate of the number of immune in the region and the number of inhabitants in Group \(G\) in the region. Let \(\pi_{rg}\) be the inclusion probability for parish \(p\) in county \(c\). An unbiased estimate of the number of immune inhabitants in region \(r\) and accumulated by group \(G\) is given by:

\[
\begin{align*}
\hat{i}_{rg} = & \sum r \sum c \sum p \sum g \frac{N_{cp,g} I_{cp,g}}{\pi_{op} A_{cp,g}}. \\
\end{align*}
\]

In the same way, an estimate of the number of inhabitants without antibodies is given by:

\[
\hat{s}_{rg} = \sum r \sum c \sum p \sum g \frac{N_{cp,g} (A_{cp,g} - I_{cp,g})}{\pi_{op} A_{cp,g}}.
\]

The total number of inhabitants is thus estimated by:

\[
\hat{N}_{rg} = \hat{i}_{rg} + \hat{s}_{rg} = \sum r \sum c \sum p \sum g \frac{N_{cp,g}}{\pi_{op}}.
\]

The number of inhabitants can, of course, be obtained from official statistics. Statistical experience and theory imply that a ratio estimate often has a greater precision if positively correlated estimates are used in the nominator and denominator.\(^1\) For this reason, we shall use the estimator:

\[
\hat{\theta}_{rg} = \frac{\hat{i}_{rg}}{\hat{N}_{rg}}.
\]
It is essential to obtain an expression for the precision of these estimates. To estimate this, we have to know the inclusion probabilities for pairs of parishes. Let \( \pi_{pp'} \) be the inclusion probability for both \( p \) and \( p' \) in county \( c \).

An unbiased estimate of the variance of \( \hat{I}_{CG} \) is given by:

\[
\sum_{c \in C} \sum_{p \in P(c)} \left( \frac{\pi_{pp'} - \pi_{pp'}}{\pi_{pp'}} \right) \left( \frac{I_{CG}}{\pi_{pp'}} \right)^2 + \sum_{c \in C} \sum_{p \in P(c)} \sum_{g \in G(c)} \hat{\sigma}_{CG}^2 \pi_{pp'} \tag{4.1}
\]

where

\[
T_{CG} = \sum_{g \in G} \frac{I_{CG} N_{CG}}{A_{CG}}
\]

and \( \hat{\sigma}_{CG}^2 \) is an unbiased estimate of the variance of \( I_{CG} N_{CG}/A_{CG} \).

Similar unbiased estimators of the variances of the estimators \( \hat{S}_{CG} \) and \( \hat{N}_{CG} \) obtained by replacing \( T_{CG} \) with the corresponding sums of \( (S_{CG} - I_{CG}) N_{CG}/A_{CG} \) and \( N_{CG}/A_{CG} \) respectively.

In our calculation, we have assumed that \( I_{CG} \) is obtained as a random sample of \( A_{CG} \) individuals from the total of \( N_{CG} \). This means that we disregard the fact that the sample is systematic. With a good level of approximation, we regard \( I_{CG} \) as binomial distributed with a probability parameter equal to the proportion of immune persons in county \( c \), parish \( p \), and group \( g \). This justifies our using \( 1/A_{CG} \) as an overestimate of \( \hat{\sigma}_{CG}^2 \). However, in the following calculations, we have used \( \hat{\sigma}_{CG}^2 = \hat{\theta}_{CG} (1 - \hat{\theta}_{CG}) / A_{CG} \).

This means that we have substituted the unknown proportion of immune with the estimated for the corresponding region and accumulated group. In fact, these two ways of approximating the variance turn out to be very close. The only situation in which the difference is of any consequence is when the proportion of immune is close to 1. In this case, the first estimate of \( \sigma_{CG}^2 \) is expected to be a substantial overestimate.

If the sample is sufficiently large, the estimate, \( \hat{\theta}_{CG} \), is approximately normally distributed. Practical experience shows that a normal approximation of the distribution of the estimators is better if we transform the estimate by the logistic transformation \( \hat{\eta}_{CG} = \ln(\hat{\theta}_{CG} / (1 - \hat{\theta}_{CG})) \). Observe that

\[
\hat{\eta}_{CG} = \hat{\theta}_{CG} / (1 - \hat{\theta}_{CG}) = I_{CG} / \hat{S}_{CG}
\]

and \( \hat{\sigma}_{CG}^2 \) is the ratio of the estimated number of immune persons and the number of non-immune people in region \( r \) and accumulated group \( G \).

The asymptotic variance of this estimator can be calculated by the delta method, and confidence intervals for \( \hat{\eta}_{CG} \) can be calculated using a normal approximation. The confidence intervals for these parameters can then be transformed into confidence intervals for \( \theta_{CG} \).