Appendix A: *Results of Successive Diagnostic Tests*

**Table A1 - Diagnostic Test and Disease Status.**

<table>
<thead>
<tr>
<th>Test Result</th>
<th>Disease Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of disease (D+)</td>
<td>Absence of disease (D-)</td>
</tr>
<tr>
<td>Positive (T+)</td>
<td>A</td>
</tr>
<tr>
<td>(true positive)</td>
<td>(false positive)</td>
</tr>
<tr>
<td>Negative (T-)</td>
<td>C</td>
</tr>
<tr>
<td>(false negative)</td>
<td>(true negative)</td>
</tr>
</tbody>
</table>

Sensitivity: \( P(T+ | D+) = \frac{A}{A+C} \)
Specificity: \( P(T- | D-) = \frac{D}{B+D} \)
PPV: \( P(D+ | T+) = \frac{A}{A+B} \)

**Table A2 - Results of the Elisa Test for HIV.**

Assumptions: Sample size = 1'000'000, sensitivity = 0.95, specificity = 0.99, prevalence of the disease = 0.001.

<table>
<thead>
<tr>
<th>Test Result</th>
<th>Disease Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of disease (D+)</td>
<td>Absence of disease (D-)</td>
</tr>
<tr>
<td>Positive (T+)</td>
<td>950</td>
</tr>
<tr>
<td>Negative (T-)</td>
<td>50</td>
</tr>
</tbody>
</table>

**PPV = 0.087**

**Table A3 - Results of the Western Blot Test for HIV.**

Assumptions: Sample size = 10’940 (= individuals who tested positive with the Elisa test), sensitivity = 0.95, specificity = 0.99, prevalence = 0.087 (in those who tested positive in the Elisa test).

<table>
<thead>
<tr>
<th>Test Result</th>
<th>Disease Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of disease (D+)</td>
<td>Absence of disease (D-)</td>
</tr>
<tr>
<td>Positive (T+)</td>
<td>902</td>
</tr>
<tr>
<td>Negative (T-)</td>
<td>48</td>
</tr>
</tbody>
</table>

**PPV = 0.90**
Appendix B: Proof of Formula (4)

The proof is conducted by induction.

Formula (4) holds for \( k = 0 \) (see (2)).

Assume that it holds at the \( k^{th} \) step, \( k \geq 1 \), i.e.

\[
\text{TRP}(k) = \frac{\pi \cdot (1 - \beta)^k}{\pi \cdot (1 - \beta)^k + (1 - \pi) \cdot \alpha^k}
\]

By definition, the TRP obtained after \( k \) studies is used as the prior probability for the \((k+1)^{th}\) study, and the samples collected in either study are independent. The TRP after study \( k + 1 \), \( \text{TRP}(k+1) \), is

\[
\text{TRP}(k + 1) = \frac{\text{TRP}(k) \cdot (1 - \beta)}{\text{TRP}(k) \cdot (1 - \beta) + \alpha \cdot (1 - \text{TRP}(k))}
\]

\[
= \frac{(1 - \beta) \cdot \frac{\pi \cdot (1 - \beta)^k}{\pi \cdot (1 - \beta)^k + (1 - \pi) \cdot \alpha^k}}{(1 - \beta) \cdot \frac{\pi \cdot (1 - \beta)^k}{\pi \cdot (1 - \beta)^k + (1 - \pi) \cdot \alpha^k} + \alpha \cdot \left(1 - \frac{\pi \cdot (1 - \beta)^k}{\pi \cdot (1 - \beta)^k + (1 - \pi) \cdot \alpha^k}\right)}
\]

\[
= \frac{\pi \cdot (1 - \beta)^{k+1} \cdot \alpha^k}{\pi \cdot (1 - \beta)^{k+1} \cdot \alpha^k + (1 - \pi) \cdot \alpha^{k+1}}
\]

In a same manner, the formula for the FPRP after \( k + 1 \) studies can be proven:

\[
\text{FPRP}(k + 1) = \frac{\text{FPRP}(k) \cdot \alpha}{(1 - \text{FPRP}(k)) \cdot (1 - \beta) + \text{FPRP}(k) \cdot \alpha}
\]

\[
= \frac{(1 - \pi) \cdot \alpha^{k+1}}{\pi \cdot (1 - \beta)^{k+1} + (1 - \pi) \cdot \alpha^{k+1}} = 1 - \text{TRP}(k + 1)
\]
Appendix C: Relationship between Power and Sample Size

In a case-control association study (a similar reasoning is possible for other designs), the power can be related to the sample size under similar assumptions as those of Wacholder et al. [1]):

- The data are obtained in a case-control association study with \( N/2 \) cases and \( N/2 \) controls (balanced design).
- The null hypothesis is that the odds ratio is one (\( H_0: \text{OR}_0 = 1 \)) and the alternative hypothesis is that the odds ratio is 1.5 (\( H_1: \text{OR}_1 = 1.5 \)).
- The odds ratio is assumed to have a log-normal distribution with variance \( \sigma^2 \).
- The type I error rate \( \alpha \) is 5 percent.
- The proportion of true positives (\( q \)) is 5, 10 or 20 percent (proportion of individuals who have the disease and tested positive for the genetic variant under consideration).

The resulting relationships are summarized in Table C1.

<table>
<thead>
<tr>
<th>Test</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>( N \cdot q )</td>
<td>( n_{12} )</td>
</tr>
<tr>
<td>T-</td>
<td>( \frac{N}{2} - N \cdot q )</td>
<td>( \frac{N}{2} - n_{12} )</td>
</tr>
</tbody>
</table>

The OR depends only on \( n_{12} \). An estimate of the standard deviation of \( \ln(\text{OR}) \) is given by

\[
s = \sqrt{\frac{1}{N \cdot q} + \frac{1}{N/2 - N \cdot q} + \frac{1}{n_{12}} + \frac{1}{N/2 - n_{12}}}.
\]  
(C1)

The power of the Wald-type hypothesis test is given by

\[
1 - \beta = \Phi \left( \frac{\ln \left( \frac{\text{OR}_1}{\text{OR}_0} \right)}{\sigma / z_{1-\alpha/2}} \right),
\]  
(C2)

where \( \Phi \) is the cumulative distribution function of the standard normal distribution and \( z_{1-\alpha/2} \) is the \((1-\alpha/2)\) percentile of the standard normal distribution.
σ is numerically estimated from Table C1, where \( n_{12} \) is such that the associated p-value of \( \text{OR}_1 / \text{OR}_0 \) equals \( \alpha \), i.e. \( n_{12} \) is determined to satisfy the following equation

\[
p = \frac{2 \left( 1 - \Phi \left( \frac{\ln(\text{OR}_1 / \text{OR}_0)}{s} \right) \right)}{\alpha}, \tag{C3}
\]

where \( s \) is determined by (C1). Note that, as \( n_{12} \) is discrete, it is generally not possible to find an exact \( n_{12} \) such that the associated p-value is exactly equal to \( \alpha \). Instead, \( n_{12} \) is determined as the largest integer value among all values that lead to a p-value which is less than or equal to \( \alpha \).

As an example, assume that \( N = 500, \alpha = 0.05 \) and \( q = 0.3 \) (cf. Table C2).

**Table C2 - Distribution of Sample**

(Based on the Assumptions of Table C1; \( N=500, \alpha=0.05 \) and \( q=0.3 \)).

<table>
<thead>
<tr>
<th>Test</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>150</td>
<td>( n_{12} )</td>
</tr>
<tr>
<td>T-</td>
<td>100</td>
<td>250 - ( n_{12} )</td>
</tr>
</tbody>
</table>

By \( n_{12} \) ranging from 1 to 150 the p-values based on testing the ORs of the resulting 150 tables increase with \( n_{12} \). A portion of this relationship is represented in Figure C1.

![Figure C1 - P-value as a function of the number of controls who tested positive, \( n_{12} (N = 500, \alpha = 0.05 \) and \( q = 0.3 \).](image-url)
\( \hat{n}_{12} \) is then the threshold value determined as the maximum over the \( 1 < n_{12} < 150 \) range for which the associated p-value is less than or equal to \( \alpha = 0.05 \). As can be seen from Figure C1, \( \hat{n}_{12} = 128 \). Inserting \( \hat{n}_{12} \) into (C1) allows the estimation of \( \sigma \):

\[
\sigma = \sqrt{\frac{1}{150} + \frac{1}{100} + \frac{1}{128} + \frac{1}{122}} = 0.18
\]  

(C4)

By using the result of (C4) in (C2), one can estimate the corresponding power as being 0.61. This value can be used to estimate the TRP of this study design.

This process is iteratively repeated for all possible sample sizes (from 100 to 10000 by steps of 10), allowing to link various sample sizes to power. Finally, a polynomial curve is applied to smooth the power curve as a function of the sample size. The graph contained in Figure C2 was obtained for \( \alpha = 0.05 \) and three values of \( q \), 0.2, 0.1, and 0.05.

![Figure C2 - Power as a function of sample size for various values of q](image)

(\( \alpha = 0.05 \) and \( q = 0.05, 0.1, \) and 0.2).