

## Analytic expressions for prediction accuracies

The aim here is to obtain analytical expressions for prediction accuracies (PAs) for additive genetic contributions to susceptibility, infectivity, and recoverability. These accuracies measure correlations between the posterior means for  $\mathbf{a}_g$ ,  $\mathbf{a}_f$ , and  $\mathbf{a}_r$  (based on the data) and their true (usually unknown) values.

Here we consider a scenario that consists of  $Z$  contact groups each containing a set of  $N$  individuals with genetic relationship specified by a pedigree or genomic relationship matrix  $\mathbf{A}$ . Epidemics are instigated by means of a single infected index case. For tractability this analysis assumes that epidemic dynamics are governed by a simple SIR (or SI) model and that the infection and recovery times of all individuals are known. Furthermore, traits are assumed to be uncorrelated and the diagonals of the matrices  $\mathbf{\Omega}$  and  $\mathbf{\Psi}$ , as well as values for the transmission rate  $\beta$  and recovery rate  $\gamma$ , are taken to be exactly known (equivalent to assuming that the number of contact groups  $Z$  is large). Lastly, fixed and group effects are ignored.

Because the likelihood in Eq.(4) naturally divides into a part related to the infection process and a part related to the recovery process, so analysis can be performed separately for each. First we look at PAs for susceptibility/infectivity and later the PA for recoverability is derived.

### PAs for susceptibility/infectivity

By Bayes' theorem

$$\pi(\theta | \xi) \propto \pi(\xi | \theta) \pi(\theta) \quad (\text{A1})$$

Substituting the part corresponding to the infection process in the likelihood in Eq.(4) and the prior in Eq.(A2) in Additional file 2 leads to the posterior probability

$$\pi(\theta | \xi) \propto \prod_z \left[ \left( \prod_{j \in z} \lambda_j \right) \left( \prod_{e \in z} e^{-\Lambda_z(t_e) \times (t_e - t_{e-1})} \right) \right] e^{-\frac{1}{2} \sum_{t,m,n,u} a_{t,m} \mathbf{A}_{mn}^{-1} \mathbf{\Omega}_{tu}^{-1} a_{u,m} - \frac{1}{2} \sum_{t,m,u} \varepsilon_{t,m} \mathbf{\Psi}_{tu}^{-1} \varepsilon_{u,m}}, \quad (\text{A2})$$

where  $z$  goes over contact groups,  $j$  goes over infected individuals (the notation  $j \in z$  denotes the fact that  $j$  only goes over individuals within group  $z$  and the symbol “ $\in$ ” is subsequently used in a corresponding fashion), and  $e$  goes over infection events that occur at times  $t_e$  (not to be confused with Euler's number  $e$  also used in this expression). The quantities  $t$  and  $u$  sum over trait types (*i.e.*  $g$ ,  $f$  and  $r$  correspond, respectively, to susceptibility, infectivity and recoverability) and  $m$  and  $n$  sum over all individuals in the system.

Substituting in the explicit expression for the force of infection in Eq.(1) and using Eq.(2) gives

$$\pi(\theta | \xi) \propto \beta^{N_I} e^{-\beta J} \prod_z \left[ \prod_{j \in z} \left( e^{a_{g,j} + \varepsilon_{g,j}} \sum_{i \in j} e^{a_{f,i} + \varepsilon_{f,i}} \right) \right] e^{-\frac{1}{2} \sum_{t,m,n,u} a_{t,m} \mathbf{A}_{mn}^{-1} \mathbf{\Omega}_{tu}^{-1} a_{u,m} - \frac{1}{2} \sum_{t,m,u} \varepsilon_{t,m} \mathbf{\Psi}_{tu}^{-1} \varepsilon_{u,m}}, \quad (\text{A3})$$

where  $i$  goes over all infected individuals in group  $z$  immediately prior to the time when individual  $j$  gets infected and  $N_I$  is the total number of infected individuals over all groups. The quantity  $J$  in Eq.(A3) is given by

$$J = \sum_z \left( \sum_{e \in z} \left[ \left( \sum_{s \in e} e^{a_{g,s} + \varepsilon_{g,s}} \right) \left( \sum_{i \in e} e^{a_{f,i} + \varepsilon_{f,i}} \right) (t_e - t_{e-1}) \right] \right), \quad (\text{A4})$$

where the sums  $s$  and  $i$  go over the susceptible and infected populations in  $z$  immediately prior to event  $e$ .

The log of the posterior probability in Eq.(A3) is given by (up to a constant term)

$$l = \log[\pi(\theta | \xi)] = N_l \log(\beta) - \beta J + \sum_z \left( \sum_{j \in z} \left[ a_{g,j} + \varepsilon_{g,j} + \log \left( \sum_{i \in j} e^{a_{f,i} + \varepsilon_{f,i}} \right) \right] \right) - \frac{1}{2} \sum_{t,m,n,u} a_{t,m} A_{mn}^{-1} \Omega_{uu}^{-1} a_{u,n} - \frac{1}{2} \sum_{t,m,u} \varepsilon_{t,m} \Psi_{uu}^{-1} \varepsilon_{u,m}. \quad (\text{A5})$$

The partial derivatives of this are

$$\begin{aligned} \frac{\partial l}{\partial a_{g,m}} &= 1 - e^{a_{g,m} + \varepsilon_{g,m}} \beta \sum_{e \in m_s} \left[ \left( \sum_{i \in e} e^{a_{f,i} + \varepsilon_{f,i}} \right) (t_e - t_{e-1}) \right] - \sum_{n,u} A_{mn}^{-1} \Omega_{gu}^{-1} a_{u,n}, \\ \frac{\partial l}{\partial a_{f,m}} &= e^{a_{f,m} + \varepsilon_{f,m}} \left[ \sum_{e \in m_l} \left( \frac{1}{\sum_{i \in e} e^{a_{f,i} + \varepsilon_{f,i}}} \right) - \beta \sum_{e \in m_l} \left[ \left( \sum_{s \in e} e^{a_{g,s} + \varepsilon_{g,s}} \right) (t_e - t_{e-1}) \right] \right] \\ &\quad - \sum_{n,u} A_{mn}^{-1} \Omega_{fu}^{-1} a_{u,n}, \end{aligned} \quad (\text{A6})$$

where  $m_s$  represents those infection events that occur during the time period in which individual  $m$  is susceptible and within the contact group in which  $m$  resides. Correspondingly,  $m_l$  represents only those events during the time period in which individual  $m$  is infected.

This analytical treatment estimates the mean posterior values for the additive genetic effects for susceptibility  $\mu_g$  and infectivity  $\mu_f$  by the values that  $a_g$  and  $a_f$  take in order to maximise the posterior probability distribution in Eq.(A5). This maximum is calculated by setting the derivatives in Eq.(A6) to zero and solving. In the general case this set of equations is intractable and to make further progress we take the limit in which  $a_g$ ,  $a_f$ ,  $\varepsilon_g$ ,  $\varepsilon_f$  are all assumed to be small. Taylor expanding Eq.(A6) in these quantities we find that, to first order, the terms involving the relationship matrix dominate.

Consequently

$$\begin{aligned} 1 - \beta \sum_{e \in m_s} [I_e (t_e - t_{e-1})] - \sum_{n,u} A_{mn}^{-1} \Omega_{gu}^{-1} \mu_{u,n} &= 0, \\ \sum_{e \in m_l} \left( \frac{1}{I_e} \right) - \beta \sum_{e \in m_l} [S_e (t_e - t_{e-1})] - \sum_{n,u} A_{mn}^{-1} \Omega_{fu}^{-1} \mu_{u,n} &= 0, \end{aligned} \quad (\text{A7})$$

where  $S_e$  and  $I_e$  represent the total susceptible and infected populations immediately preceding event  $e$ . Remembering that  $\Omega$  is a diagonal matrix (*i.e.* correlations are ignored), this can be rearranged to

$$\mu_{g,n} = \Omega_{gg} \sum_m \left[ A_{nm} \left( 1 - \beta \sum_{e \in m_s} [I_e (t_e - t_{e-1})] \right) \right] \quad (\text{A8})$$

and

$$\mu_{f,n} = \Omega_{ff} \sum_m \left[ A_{nm} \left( \sum_{e \in m_I} \left( \frac{1}{I_e} \right) - \beta \sum_{e \in m_I} [S_e(t_e - t_{e-1})] \right) \right]. \quad (\text{A9})$$

These two expressions are estimated in the following two subsections:

### Estimating $\mu_{g,n}$

Suppose we simulate from the model using a given set of model parameters  $\theta$ . The various quantities in Eq.(A8) can be estimated. First we start with

$$\sum_{e \in m_S} [I_e(t_e - t_{e-1})], \quad (\text{A10})$$

where the sum goes over events in which individual  $m$  is susceptible. For brevity this quantity is denoted by  $\kappa_m$ . The values of  $\kappa_m$  vary drastically between individuals because some become infected earlier on within the epidemic and some later. Furthermore, the order in which individuals becomes infected is essentially random, so to simplify matters we replace Eq.(1) with its mean field approximation

$$\lambda_n = \beta e^{a_{g,n} + \varepsilon_{g,n}} \overline{e^{a_f + \varepsilon_f}}^z I, \quad (\text{A11})$$

where  $I$  represents the number of currently infected individuals and the bar with superscript  $z$  indicates that the quantity is averaged over all individuals within group  $z$ .

Focusing on individual  $m$ , the distribution in the likelihood from Eq.(A3) is given by

$$\pi(\kappa_m | \theta, \xi_{-m}) \propto e^{-\beta e^{a_{g,m} + \varepsilon_{g,m}} \overline{e^{a_f + \varepsilon_f}}^z \kappa_m}. \quad (\text{A12})$$

This implies that  $\kappa_m$  is, in fact, exponentially distributed, a result which can be expressed as

$$\kappa_m = \frac{u_m}{\beta e^{a_{g,m} + \varepsilon_{g,m}} \overline{e^{a_f + \varepsilon_f}}^z}, \quad (\text{A13})$$

where  $u_m$  is an exponentially distributed random variable with unit mean. An intuitive reason for this expression is that if an individual is more susceptible than the population as a whole (*i.e.*  $a_{g,m} + \varepsilon_{g,m}$  is positive) then on average it will tend to get infected earlier on within the epidemic (subject to large stochastic fluctuations coming from  $u_m$ ) and hence  $\kappa_m$  is expected to typically be smaller. Similarly, if the contact group contains many highly infectious individuals  $\overline{e^{a_f + \varepsilon_f}}^z$  will be significantly larger than one meaning that epidemics proceed more quickly, again reducing  $\kappa_m$ .

Substitution of Eq.(A13) into Eq.(A8) leads to (assuming effect sizes are small such that the exponential terms can be Taylor expanded)

$$\mu_{g,n} = \Omega_{gg} \sum_m A_{nm} \left( 1 - \left[ 1 - a_{g,m} - \varepsilon_{g,m} - \overline{a_f}^z \right] u_m \right). \quad (\text{A14})$$

For future manipulation it is of use to rearrange this as

$$\mu_{g,n} = \Omega_{gg} \sum_z \left[ \sum_{j \in z} A_{nj} \left( 1 - \left[ 1 - a_{g,j} - \varepsilon_{g,j} - \frac{1}{N} \sum_{q \in z} (a_{f,q} + \varepsilon_{f,q}) \right] u_j \right) \right], \quad (\text{A15})$$

where  $z$  sums over contact groups and  $j$  and  $q$  sum over all individuals within group  $z$ .

### Estimating $\mu_{f,n}$

Equation (A9) can be rewritten as

$$\mu_{f,n} = \Omega_{ff} \sum_z \left[ \sum_{e \in z} \left( \sum_{j \in e} A_{nj} \left( \frac{1}{I_e} - \beta S_e (t_e - t_{e-1}) \right) \right) \right], \quad (\text{A16})$$

where  $z$  goes over contact groups,  $e$  goes over all infection events (after the initially infected index case), and  $j$  sums over all infected individuals immediately prior to infection event  $e$ .

The time difference between consecutive events is exponentially distributed (because events are Markovian) according to

$$t_e - t_{e-1} = \frac{u_e}{\beta \left( \sum_{s \in e} e^{a_{g,s} + \varepsilon_{g,s}} \right) \left( \sum_{i \in e} e^{a_{f,i} + \varepsilon_{f,i}} \right)}, \quad (\text{A17})$$

where  $s$  goes over all susceptible individuals immediate prior to time  $t_e$  and  $i$  goes over all infected individuals. The random variables  $u_e$  are exponentially distributed with a mean of one. Substituting Eq.(A17) into Eq.(A16), as well as assuming that phenotypic effects are small, leads to

$$\mu_{f,n} = \Omega_{ff} \sum_z \left[ \sum_{e=1}^{N-1} \left( \sum_{j \in e} A_{nj} \left( \frac{1}{e} - \frac{u_e}{e} \left[ 1 - \frac{1}{N-e} \sum_{s \in e} (a_{g,s} + \varepsilon_{g,s}) - \frac{1}{e} \sum_{i \in e} (a_{f,i} + \varepsilon_{f,i}) \right] \right) \right) \right], \quad (\text{A18})$$

Note, here events have been numbered such that  $e=1$  corresponds to the first infection in group  $z$  after the index case all the way up to  $e=N-1$ , where  $N$  is the number of individuals within each contact group. This expression assumes a simple SI model dynamics for the system, *i.e.* recoveries are ignored. This is justified by the fact that most of the information regarding an individual's infectivity comes from when it has just become infected (considering the begin of an epidemic the index case definitely infects the second individual, then one of the first two cases infects the third etc... successively less information is available regarding who is infecting whom).

### Expressions for PAs

We now estimate the susceptibility and infectivity PAs for a given individual  $n$ . This is defined by the correlation between its estimated and true value:

$$\alpha_{g,n} = \frac{\langle \mu_{g,n} a_{g,n} \rangle - \langle \mu_{g,n} \rangle \langle a_{g,n} \rangle}{\sqrt{\left( \langle a_{g,n}^2 \rangle - \langle a_{g,n} \rangle^2 \right) \left( \langle \mu_{g,n}^2 \rangle - \langle \mu_{g,n} \rangle^2 \right)}}, \quad (\text{A19})$$

$$\alpha_{f,n} = \frac{\langle \mu_{f,n} a_{g,n} \rangle - \langle \mu_{f,n} \rangle \langle a_{f,n} \rangle}{\sqrt{\left( \langle a_{f,n}^2 \rangle - \langle a_{f,n} \rangle^2 \right) \left( \langle \mu_{f,n}^2 \rangle - \langle \mu_{f,n} \rangle^2 \right)}}.$$

where the angular brackets represent averages over potential realisations from the system (e.g., this is equivalent to simulating from the system a large number of times and taking the average over those simulations). Note, unlike alternative definitions which rely of averages taken over the entire population (or sub-population, such as sires), this allows us to generate PAs for each individual (which makes sense because those individuals with more relatives are necessarily going to have higher PAs).

The means of quantities are zero  $\langle a_{g,n} \rangle = \langle a_{f,n} \rangle = \langle \mu_{g,n} \rangle = \langle \mu_{f,n} \rangle = 0$  and the variances of the additive genetic effects are given by the covariance matrix  $\mathbf{\Omega}$ . Consequently Eq.(A19) simplifies to

$$\alpha_{g,n} = \frac{\langle \mu_{g,n} a_{g,n} \rangle}{\sqrt{\Omega_{gg} \langle \mu_{g,n}^2 \rangle}}, \quad \alpha_{f,n} = \frac{\langle \mu_{f,n} a_{f,n} \rangle}{\sqrt{\Omega_{ff} \langle \mu_{f,n}^2 \rangle}}. \quad (\text{A20})$$

Based on the expressions in Eq.(A15) and Eq.(A18), the follow quantities can be identified:

$$\begin{aligned} \langle \mu_{g,n} a_{g,n} \rangle &= \Omega_{gg} \sum_z \left[ \sum_{j \in z} A_{nj}^2 \right], \\ \langle \mu_{f,n} a_{f,n} \rangle &= \Omega_{ff} \sum_z \left[ \frac{1}{N} \sum_{j,q \in z} A_{nj} A_{nq} v_{jq} \right], \\ \langle \mu_{g,n}^2 \rangle &= \sum_z \left[ \sum_{j \in z} A_{nj}^2 \right] + \sum_{z,z'} \left[ \sum_{j \in z, j' \in z'} A_{nj} A_{nj'} \left( \Omega_{gg} A_{jj'} + \Psi_{gg} I_{jj'} \right. \right. \\ &\quad \left. \left. + \frac{1}{N^2} \sum_{q \in z, q' \in z'} (\Omega_{ff} A_{qq'} + \Psi_{ff} I_{qq'}) \right) \right], \\ \langle \mu_{f,n}^2 \rangle &= \sum_z \left[ \frac{1}{N} \sum_{j \in z} v_{jj} A_{nj}^2 \right] + \sum_{z,z'} \left[ \frac{1}{N^2} \sum_{j,q \in z, j',q' \in z'} A_{nj} A_{nj'} \right. \\ &\quad \left. \times \left( \omega_{jqj'q'} \left( \Omega_{gg} A_{qq'} + \Psi_{gg} I_{qq'} \right) + \eta_{jqj'q'} \left( \Omega_{ff} A_{qq'} + \Psi_{ff} I_{qq'} \right) \right) \right]. \end{aligned} \quad (\text{A21})$$

Here  $z$  sums over contact groups and  $j$  and  $q$  sum over all individuals within  $z$ . Similarly  $z'$  sums over contact groups and  $j'$  and  $q'$  sum over all individuals within  $z'$ . Finally,  $\mathbf{A}$  is the relationship matrix and  $\mathbf{I}$  is the identity matrix.

These expression involve the following tensors:

$$\begin{aligned} \tau_{jq} &= \left\langle \sum_{e=E_j}^{E_q-1} \frac{N}{e(N-e)} \right\rangle, \quad v_{jq} = \left\langle \sum_{e=\max(E_j, E_q)}^{N-1} \frac{N}{e^2} \right\rangle, \\ \omega_{jqj'q'} &= \left\langle \left( \sum_{e=E_j}^{E_q-1} \frac{N}{e(N-e)} \right) \left( \sum_{e'=E_{j'}}^{E_{q'}-1} \frac{N}{e'(N-e')} \right) \right\rangle, \\ \eta_{jqj'q'} &= \left\langle \left( \sum_{e=\max(E_j, E_q)}^{N-1} \frac{N}{e^2} \right) \left( \sum_{e'=\max(E_{j'}, E_{q'})}^{N-1} \frac{N}{e'^2} \right) \right\rangle, \end{aligned} \quad (\text{A22})$$

where  $E_j$  is the value of  $e$  for which individual  $j$  becomes infected and the angle brackets average over random orderings for  $E_j$ .

To make progress the expressions in Eq.(A22) are replaced by integral approximations. For example

$$\begin{aligned}
v_{jj} &= \left\langle \sum_{e=E_j}^{N-1} \frac{N}{e^2} \right\rangle \cong \frac{1}{N-1} \int_{\frac{1}{2}}^{N-\frac{1}{2}} \int_{e=E_j}^N \frac{N}{e^2} de dE_j, \\
&\cong \frac{1}{N-1} \int_{\frac{1}{2}}^{N-\frac{1}{2}} \left( \frac{N}{E_j} - 1 \right) dE_j, \\
&\cong \frac{N \log(2N-1) - (N-1)}{N-1}, \\
&\cong \log(2N).
\end{aligned} \tag{A23}$$

The last line comes from assuming that  $N$  is reasonably large. Integral approximations for all the other quantities in Eq.(A22) yield the following results:

$$\begin{aligned}
\tau_{jq} &= \begin{cases} 1 & \text{if } j \neq q \\ 0 & \text{if } j = q \end{cases}, & v_{jq} &= \begin{cases} 1 & \text{if } j \neq q \\ \log(2N) & \text{if } j = q \end{cases}, \\
\omega_{jqj'q'} &= \begin{cases} \tau_{jq}\tau_{j'q'} & \text{if } j \text{ and } q \text{ different from } j' \text{ and } q' \\ 3 & \text{if } j = j' \neq q = q' \\ 2 & \text{if } j = j' \neq q \neq q' \text{ or } q = q' \neq j \neq j' \\ 0 & \text{if } j = q' \neq j' = q \end{cases}, \\
\eta_{jqj'q'} &= \begin{cases} v_{jq}v_{j'q'} & \text{if } j \text{ and } q \text{ different from } j' \text{ and } q' \\ 3.5 & \text{if } j = q = j' = q' \\ 2 & \text{if } j = j' \neq q \neq q' \text{ or } j = q' \neq q \neq j' \\ & \text{or } q = j' \neq j \neq q' \text{ or } q = q' \neq j \neq j' \\ \log^2(N) & \text{if } j \neq j' = q = q' \text{ or } j' \neq q = q' = j \\ & \text{or } q \neq j = j' = q' \text{ or } q' \neq q = j = j' \\ 2\log(N) & \text{if } j = j' \neq q = q' \text{ or } j = q' \neq q = j' \end{cases}
\end{aligned} \tag{A24}$$

Substituting the expressions from Eq.(A21) into Eq.(A20) gives the final results:

$$\begin{aligned}
\alpha_{g,n} &= \frac{\sqrt{\Omega_{gg}} \sum_s A_{ns}^2}{\sqrt{\sum_s A_{ns}^2 + \sum_{z,z'} \left[ \sum_{j \in z, j' \in z'} A_{nj} A_{nj'} \left( \Omega_{gg} A_{jj'} + \Psi_{gg} \mathbf{I}_{jj'} \right. \right. \\
&\quad \left. \left. + \frac{1}{N^2} \sum_{q \in z, q' \in z'} \left( \Omega_{ff} A_{qq'} + \Psi_{ff} \mathbf{I}_{qq'} \right) \right] \right]}, \\
\alpha_{f,n} &= \frac{\sqrt{\Omega_{ff}} \left( \frac{\log(2N)}{N} \sum_m A_{nm}^2 + \sum_z \left[ \frac{1}{N} \sum_{j \neq q \in z} A_{nj} A_{nq} \right] \right)}{\sqrt{\frac{\log(2N)}{N} \sum_m A_{nm}^2 + \sum_{z,z'} \left[ \frac{1}{N^2} \sum_{j,q \in z, j',q' \in z'} A_{nj} A_{nj'} \left( \omega_{jqj'q'} \left( \Omega_{gg} A_{qq'} + \Psi_{gg} \mathbf{I}_{qq'} \right) \right. \right. \\
&\quad \left. \left. + \eta_{jqj'q'} \left( \Omega_{ff} A_{qq'} + \Psi_{ff} \mathbf{I}_{qq'} \right) \right] \right]}}.
\end{aligned} \tag{A25}$$

To reiterate from above:  $z$  sums over contact groups and  $j$  and  $q$  sum over all individuals within  $z$ ,  $z'$  also sums over contact groups and  $j'$  and  $q'$  sum over all individuals within  $z'$ , and  $N$  is the number individuals within each contact group.

## PA for recoverability

Substituting the part corresponding to the recovery process in the likelihood Eq.(4) and the prior in Additional file 2 Eq.(A2) into Eq.(A1) leads to the posterior probability

$$\pi(\theta | \xi) \propto \left( \prod_p F_\Gamma(\delta t_p | w_p, k) \right) \times e^{-\frac{1}{2\Omega_{rr}} \sum_{m,n} a_{r,m} A_{mn}^{-1} a_{r,n} - \frac{1}{2\Psi_{rr}} \sum_m \varepsilon_{r,m}^2}, \quad (\text{A26})$$

where  $p$  sums over all individual that get infected (including the index cases), and  $m$  and  $n$  sum over all individuals (whether they take part in epidemics or not). The term

$$F_\Gamma(\delta t_p | w_p, k) = \Gamma(k)^{-1} w_p^{-k} k^k \delta t_p^{k-1} e^{-k\delta t_p/w_p} \quad (\text{A27})$$

is the gamma distribution, assuming that individuals have a mean infection duration  $w_p$ , as defined in Eq.(1). Substituting Eq.(1) into Eq.(A27) and taking the logarithm gives

$$l = \log[\pi(\theta | \xi)] = \sum_p k (a_{r,p} + \varepsilon_{r,p}) + (k-1) \log(\delta t_p) - k\gamma \delta t_p e^{a_{r,p} + \varepsilon_{r,p}} - \frac{1}{2\Omega_{rr}} \sum_{m,n} a_{r,m} A_{mn}^{-1} a_{r,n} - \frac{1}{2\Psi_{rr}} \sum_m \varepsilon_m^2. \quad (\text{A28})$$

The partial derivatives of this are

$$\frac{\partial l}{\partial a_{g,m}} = k \left( 1 - \gamma \delta t_m e^{a_{r,m} + \varepsilon_{r,p}} \right) - \frac{1}{\Omega_{rr}} \sum_n A_{mn}^{-1} a_{r,n} \quad (\text{A29})$$

This analytical treatment estimates the mean posterior values for the additive genetic effects for the recoverability  $\mu_r$  by the values that  $a_r$  take in order to maximise the posterior probability distribution in Eq.(A28). This maximum is calculated by setting the derivatives in Eq.(A29) to zero and solving. In the general case this set of equations is intractable and to make further progress we take the limit in which  $a_r$  and  $\varepsilon_r$  are assumed to be small. Taylor expanding Eq.(A29) in these quantities we find that to first order the terms involving the relationship matrix dominate. This gives

$$\mu_{r,n} = k\Omega_{rr} \sum_m \left[ A_{nm} (1 - \gamma \delta t_m) \right]. \quad (\text{A30})$$

The recovery times can be written as

$$\delta t_m = \frac{u_m}{\gamma e^{a_{r,m} + \varepsilon_{r,m}}}, \quad (\text{A31})$$

where  $u_m$  is a gamma distributed variable with mean 1. Substituting this into Eq.(A30) and making use of the fact that  $a_{r,m}$  and  $\varepsilon_{r,m}$  are small leads to

$$\mu_{r,n} = k\Omega_{rr} \sum_m \left[ A_{nm} \left( 1 - \left[ 1 - a_{r,m} - \varepsilon_{r,m} \right] u_m \right) \right]. \quad (\text{A32})$$

The following quantities can be derived from this equation:

$$\begin{aligned}
\langle \mu_{r,n} a_{r,n} \rangle &= \Omega_{rr} \sum_m A_{nm}^2, \\
\langle \mu_{r,n}^2 \rangle &= \frac{1}{k} \sum_m A_{nm}^2 + \sum_{mq} A_{nm} A_{nq} (\Omega_{rr} A_{mq} + \Psi_{rr} I_{mq}),
\end{aligned}
\tag{A33}$$

where  $\mathbf{A}$  is the relationship matrix,  $\mathbf{I}$  is the identity matrix and  $k$  is the shape parameter. From these quantities the PA can be derived:

$$\alpha_{r,n} = \frac{\langle \mu_{r,n} a_{r,n} \rangle}{\sqrt{\Omega_{rr} \langle \mu_{r,n}^2 \rangle}} = \frac{\sqrt{\Omega_{rr} \sum_m A_{nm}^2}}{\sqrt{\frac{1}{k} \sum_m A_{nm}^2 + \sum_{mq} A_{nm} A_{nq} (\Omega_{rr} A_{mq} + \Psi_{rr} I_{mq})}}.
\tag{A34}$$