Using pooled data to estimate genetic parameters for social interaction traits

Katrijn Peeters, Esther Ellen and Piter Bijma
Pooling phenotypic data

Collecting individual data in group housing systems

Expensive
Difficult

Pooling data
Social interaction theory

\[ P_i = A_{D_i} + E_{D_i} \]
Social interactions theory

\[ P_i = A_{D_i} + E_{D_i} + \sum_{i \neq j}^{n-1} A_{S_j} + \sum_{i \neq j}^{n-1} E_{S_j} \]
Direct model $\Leftrightarrow$ Direct-Social model

Direct

\[ A_D \]
\[ \sigma_{A_D}^2 \]

Direct-indirect

\[ A_T = A_D + 3A_S \]
\[ \sigma_{A_T}^2 = \sigma_{A_D}^2 + 6\sigma_{A_{DS}} + 9\sigma_{A_S}^2 \]
Aim: what can we estimate?

Individual data

$A_D$  $A_S$

$A_T$

Pooled data

$A_D$  $A_S$

$A_T$  $?$
Three Methods

Simulated data
Theoretical derivation
Empirical data

Can we estimate direct and social genetic parameters from pooled data?
1. Theoretical derivation: Individual data

\[ P_1 = A_{D1} + A_{S2} + A_{S3} + A_{S4} + E \]

\[ P_2 = A_{D2} + A_{S1} + A_{S3} + A_{S4} + E \]

\[ P_3 = A_{D3} + A_{S1} + A_{S2} + A_{S4} + E \]

\[ P_4 = A_{D4} + A_{S1} + A_{S2} + A_{S3} + E \]

Information on \( A_D \) and \( A_S \) is obtained from different phenotypes (\( P_1 \neq P_{2/3/4} \))

Both \( A_D \) and \( A_S \) are estimable
1. Theoretical derivation: Pooled data ($P^*$)

$P^* = A_D1 + A_D2 + A_D3 + A_D4 + 3(A_S1 + A_S2 + A_S3 + A_S4) + E$

An individual’s $A_D$ and its full $A_S$ is always lumped into a single record ($P^*$)

$\Rightarrow A_D$ and $A_S$ are fully confounded

$P^* = A_T1 + A_T2 + A_T3 + A_T4 + E$

$\Rightarrow A_T$ can be estimated

$\Rightarrow y = Xb + \Sigma Za + e$ will estimate $A_T$
2. Simulated data

- 500 Sires, 500 Dams, 12 offspring per mating
- 12,000 individuals
- Direct and Social effects
- 3,000 groups of 4 individuals each
- Record = pooled phenotype per group
2. Simulated data

- **Input**
  \[
  \sigma_{AD}^2 = 1; \; \sigma_{AS}^2 = 1; \; \sigma_{ADS} = 0.5 \\
  \sigma_{ED}^2 = 2; \; \sigma_{ES}^2 = 2; \; \sigma_{EDS} = 0
  \]

- **Pooled data model (ASReml)**
  \[
y = \mathbf{Xb} + \sum \mathbf{Za} + \mathbf{e}
  \]

  \[y \sim \mu \; \text{FIXED } \; !r \; A_1 \; \text{and}(A_2) \; \text{and}(A_3) \; \text{and}(A_4)\]

- **Expected values ⇔ obtained values (100 replicates)**
  \[
  \sigma_{AT}^2 = \sigma_{AD}^2 + 6 \; \sigma_{ADS} + 9 \; \sigma_{AS}^2 = 13 \; \Leftrightarrow \; \hat{\sigma}_A^2 = 13.05 (\pm 2.12) \\
  \sigma_{E^*}^2 = 4(\sigma_{ED}^2 + 6 \; \sigma_{EDS} + 9 \; \sigma_{ES}^2) = 80 \; \Leftrightarrow \; \hat{\sigma}_E^2 = 80.32 (\pm 7.30)
  \]

  Note: \( \hat{\sigma}_A^2 \neq \hat{\sigma}_{AD}^2 \) but \( \hat{\sigma}_A^2 = \hat{\sigma}_{AT}^2 \)
3. Empirical data: cannibalistic laying hens

- 12,944 Laying hens
  - 6,092 W1
  - 6,852 WB

- Individual survival time:
  \[ y \sim \mu \text{ FIXED } ! r \ A_{D1} \ A_{I2} \text{ and } (A_{I3}) \text{ and } (A_{I4}) \text{ Cage} \]

- Group survival time (pooled per cage):
  \[ y \sim \mu \text{ FIXED } ! r \ A_{T1} \text{ and } (A_{T2}) \text{ and } (A_{T3}) \text{ and } (A_{T4}) \]
3. Empirical data

### Individual data

<table>
<thead>
<tr>
<th></th>
<th>W1</th>
<th>WB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{AD}^2$</td>
<td>705 (±171)</td>
<td>1,404 (±301)</td>
</tr>
<tr>
<td>$\sigma_{ADS}$</td>
<td>59 (±61)</td>
<td>-162 (±105)</td>
</tr>
<tr>
<td>$\sigma_{AS}^2$</td>
<td>104 (±41)</td>
<td>292 (±72)</td>
</tr>
<tr>
<td>$\sigma_{Cage}^2$</td>
<td>799 (±166)</td>
<td>1,191 (±238)</td>
</tr>
<tr>
<td>$\sigma_E^2$</td>
<td>7,980 (±210)</td>
<td>12,675 (±365)</td>
</tr>
</tbody>
</table>

### Pooled data

\[
\sigma_{AD}^2 + 6 \sigma_{ADS} + 9 \sigma_{AS}^2 = \sigma_{AT}^2
\]

\[
16 \sigma_{Cage}^2 + 4 \sigma_E^2 = \sigma_{E^*}^2
\]
3. Empirical data

### Individual data

<table>
<thead>
<tr>
<th></th>
<th>W1</th>
<th>WB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_{AT}$</td>
<td>1,996 (±640)</td>
<td>2,521 (±842)</td>
</tr>
<tr>
<td>$\sigma^2_{E*}$</td>
<td>44,700 (±2,526)</td>
<td>69,752 (±3,513)</td>
</tr>
</tbody>
</table>

### Pooled data

<table>
<thead>
<tr>
<th></th>
<th>W1</th>
<th>WB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_{AT}$</td>
<td>1,979 (±643)</td>
<td>2,521 (±845)</td>
</tr>
<tr>
<td>$\sigma^2_{E*}$</td>
<td>44,750 (±2,538)</td>
<td>69,750 (±3,519)</td>
</tr>
</tbody>
</table>

Indeed: analysis of pooled records yields $\hat{\sigma}^2_{AT}$ rather than $\hat{\sigma}^2_{AD}$
3. Empirical data: The cost of pooling

Pooling causes surprisingly little reduction in response to selection for socially-affected traits.
Conclusion

With social interactions, estimated $\text{Var}(A)$ from pooled data will differ from ordinary $\text{Var}(A)$.

$$\text{Var}(A)_{\text{pooled}} = \text{Var}(A_T)$$
Social interactions

Peeters et al., Genetics Selection Evolution 45:27
katrijn.peeters@wur.nl

Using pooled data to estimate variance components and breeding values for traits affected by social interactions

Katrijn Peeters*, Esther Dorien Ellen and Piter Bijma