A Tale of Two Parasites
Statistical modelling to support NTD control in Africa
Peter J Diggle
Lancaster University
MDA: a tool for control of vector-borne filarial disease

- Ivermectin (Mectizan): annual dose clears microfilarial infections of the blood
- generally considered safe, with no serious side-effects
- mass distribution made possible by donation programme (Merck)
- used in multi-national programmes to combat onchocerciasis and lymphatic filariasis
Onchocerciasis: distribution
The Loa loa problem

People who are heavily co-infected with *Loa loa* parasites can experience serious (occasionally fatal) adverse reactions to ivermectin
Individuals at risk of experiencing an SAE are those with high microfilarial loads.

Difficult/expensive to measure individual MF load.

Individuals with high MF loads are likely to be found in high-prevalence areas.

High-risk area $\iff$ Prevalence $> 20\%$
“The answer to any prediction problem is a probability distribution”

Peter McCullagh

\[ S = \text{state of nature} \]
\[ Y = \text{all relevant data} \]
\[ T = \mathcal{F}(S) = \text{target for prediction} \]

Model: \[ [S, Y] = [S][Y|S] \]
Prediction: \[ [S, Y] \Rightarrow [S|Y] \Rightarrow [T|Y] \]

\[ T(x) = \text{prevalence} > 20\% \quad \text{(yes/no)} \]
Figure 6: PCM for [high risk] in Cameroon based on ERMr with ground truth data.
Second solution... RAPLOA
Calibration surveys

- 223 villages
- 24 to 229 individuals per village, total 19,128
Calibration data
RAPLOA survey locations
RAPLOA exceedance map
Third solution... use prevalence to predict numbers of high-risk individuals

- Calibration data include individual MF loads
- Wide variation in prevalence and MF load
- But distributional shape broadly consistent across all villages
Schematic: $P =$ prevalence; $T =$ proportion highly infected
Target for prediction: \( T = \rho(U) \times \{1 - G(20000; \lambda(V), \kappa)\} \)
Fourth solution... the Loascope

Mobile Microscopy
taking imaging to new places
Comparing traditional microscopy and Loascope

15,181 people from 88 villages in Cameroon, each tested by traditional microscopy (MF) and Loascope (LS).

- High concordance wrt infection status ($\kappa = 0.858$)

<table>
<thead>
<tr>
<th></th>
<th>CS 0</th>
<th>CS 1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MF 0</td>
<td>12135</td>
<td>357</td>
<td>12492</td>
</tr>
<tr>
<td>MF 1</td>
<td>268</td>
<td>2421</td>
<td>2689</td>
</tr>
<tr>
<td>Total</td>
<td>12413</td>
<td>2779</td>
<td>15181</td>
</tr>
</tbody>
</table>

- Ranges of discordant values:
  - $MF = 0 \Rightarrow 31 \leq LS \leq 7261$
  - $LS = 0 \Rightarrow 20 \leq MF \leq 19940$

- Parasite-loads well-calibrated except at very high MF loads
Variation between villages: exploiting correlation

- The proportion of a rare condition (here, $MF > 20K$) in a sample is an imprecise estimate of the proportion in the population

- Correlation between community-level prevalence and intensity of infection ⇒ information gain (narrower prediction intervals)

Both prevalence and intensity give important information and both are necessary for efficient prediction
Model-validation study in Cameroon

1. Loascope data from approx 100 individuals in each of 30 villages

2. Use model-based decision rule:
   - $P = \text{probability that } < 1\% \text{ of individuals have } > 20000 \text{ parasites/ml}$
   - MDA indicated if $P > 0.95$

3. Additional sample/census in each village to confirm (or not) model-based decision
Summary Results

Phase 1: Results based on algorithm predications

<table>
<thead>
<tr>
<th></th>
<th>MDA</th>
<th>TNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDA</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>TNT</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><strong>14</strong></td>
<td><strong>10</strong></td>
</tr>
</tbody>
</table>

Phase 2: Validation

Algorithm-based predictions are conservative, resulting in 5 villages receiving TNT that *could* have received MDA, based on validation results.
Summary Results

<table>
<thead>
<tr>
<th>Phase 2: Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDA</td>
</tr>
<tr>
<td>MDA 14</td>
</tr>
<tr>
<td>TNT 0</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

No villages deemed unsafe to treat with MDA in validation phase were targeted for MDA based on algorithm predictions.
- Model-based strategy converges to test-and-not-treat as appetite for risk decreases

- Zero risk is unattainable
  - 20000/ml a uniquely correct threshold?
  - measurement error in estimates of parasites/ml?

- Sequential algorithm:
  
  Initial sample $\Rightarrow P(\text{“safe” for MDA}) = p$
  
  $p > 0.95$  \text{ “safe”}
  
  $p < 0.05$  \text{ “not safe”}
  
  $0.05 < p < 0.95$  \text{ increase sample size}
The geospatial extension

U(x): log-odds prevalence

V(x): log-intensity

⇒ narrower prediction intervals (borrowing strength)
Mapping areas “safe” for MDA

“safe” ⇔ “at least 95% certain that at most 1% of individuals have more than 20,000 parasites/ml
Fifth solution...what’s the real question?

- **Vaccine safety analogue** \( P(SAE) = ? \)
- **Loa loa...what do we know?**
  - **Predictive distribution** of \( Y = MF \) load at location \( x \): \( f_x(y | \text{data}) \)
  - **Risk of SAE** given \( MF = y \)
    \[
p(y) = \Phi \{-1.05 + (1.87[\log_e(y) - 11.06])\} \pm ?
    \]
    Boussinesq et al (2003, corrected)
- **Predictive target**
  \[
  T(x) = \text{Prob}(\text{SAE at } x) = \int p(y)f_x(y | \text{data})dy
  \]
- **Draw samples** from predictive distribution of \( T(x) \) over required area
Conclusions

- Principled statistical methods
  - force assumptions into the open
  - deliver optimal solutions within the declared assumptions

- But there is no free lunch

  "We buy information with assumptions"

  C H Coombs

- analyse problems, not data

  "Better an approximate answer to the right question than a precise answer to the wrong question"

  John Tukey

