Introduction

• What is “Qualitative Analysis”?  
• Characterising uncertainty and method performance 
• Qualitative response dependent on a concentration 
• What can we expect from labs?
What is qualitative analysis?

- “Classification according to specific criteria”
  - “Above” or “Below” a limit
  - “Within Spec.”
  - “Red”
  - Classification into ranges (<2; 2-5; 5-10; >10)
  - Molecular species by NMR, IR, MS…..
  - Material or ingredient (“Rubber”, “Fat”…)
  - Origin or authenticity

Expression of uncertainty in qualitative analysis

- False response rates
  - What is a false response rate?
  - How is it determined?
- Alternative expressions of method performance or uncertainty
- Logistic regression for modelling performance

NOTE
Current literature refers to “nominal properties”
### False response rates and derived indicators

<table>
<thead>
<tr>
<th>Actual (True) value</th>
<th>Negative</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>TN</td>
<td>FN</td>
</tr>
<tr>
<td>Positive</td>
<td>FP</td>
<td>TP</td>
</tr>
</tbody>
</table>

#### Observed

<table>
<thead>
<tr>
<th>Reliability Measure</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>False positive rate</td>
<td>$\frac{FP}{(TN + FP)}$</td>
</tr>
<tr>
<td>False negative rate</td>
<td>$\frac{FN}{(TP + FN)}$</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>$\frac{TP}{(TP + FN)}$</td>
</tr>
<tr>
<td>Specificity</td>
<td>$\frac{TN}{(TN + FP)}$</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>$\frac{TP}{(TP + FP)}$</td>
</tr>
<tr>
<td>Efficiency</td>
<td>$\frac{(TP + TN)}{(TP + TN + FP + FN)}$</td>
</tr>
<tr>
<td>Youden Index</td>
<td>$\text{Sensitivity} + \text{Specificity} - 100$</td>
</tr>
<tr>
<td>Likelihood ratio</td>
<td>$\frac{(1-\text{False neg. rate})}{\text{False pos. rate}}$</td>
</tr>
<tr>
<td>Bayes posterior probability</td>
<td>Bayes rule (requires ‘prior’) + valuable for cumulative data</td>
</tr>
</tbody>
</table>

### Alternative performance indicators (Single laboratory)

#### Proportion of positives that are correct

#### Uncertainty about the result
False response rates - how much data?

- Observed: 7/126 (5.6%)
- 95% confidence interval (binomial)
  - 1.6% to 9.5%
- 95% CI proportional to \(1/\sqrt{n_{\text{obs}}}\)
  - needs a LOT of false responses for precise figures
  - but false responses are rare for good methods….
- Most useful direct studies are ‘worst case’ or near 50% false response levels

False responses: Estimation from thresholds

\[
x_1 = 0 \quad \text{threshold} \quad x_2 > 0
\]

- "negative" region
- "positive" region
- Concentration
False responses: From probabilities

• Spectroscopic identification study
• Calculated chance FT-IR match probabilities
  • probabilities based on “match-binning” - hits within set distance
  • required hypergeometric distribution ($n$ matches of $m$ taken from population)
• Compared with actual hits on IR database

False responses: From probabilities

• Theoretical predictions very sensitive to probability assumptions
  • 10% changes in $p$ make large differences in predictions
• Best performance within factor of 3-10
  • (Improved over binomial probabilities by >10$^6$)
• Probability information must be excellent for good predictions
False response rates from databases

• Most spectral databases contain 1 of each material
• most populations do not!
• Population data must account for sub-populations
• cf. DNA profiles for racially inhomogeneous populations

Using Logistic Regression

• Logistic regression models probability as a function of a continuous variable

\[
E(Y \mid x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}
\]

• Example:
• p(DNA found) vs DNA concentration
Logistic regression and performance assessment

DNA detection study
- 6 x 6 replicates each

Logistic regression and method performance

Amylase detection vs % pasteurised
- 13 laboratories
- 8 samples
- 2 replicates each
Logistic regression and method performance

Problems for qualitative “uncertainty”

- Hard to estimate low false response rates
  - May take hundreds of experiments
- Harder to estimate population probabilities
- Harder still to evaluate joint probabilities
  - … and these have large effects on calculation
- Prior probabilities are very rarely available
Recommendations*

• It is realistic to expect that testing laboratories have qualitative test method parameters (conditions of testing) under adequate control. Evidence will typically involve
  – evidence of traceability for the values of important control parameters prescribed by the method
  – evidence that uncertainties in these parameters are sufficiently small for the purpose

• It is important for laboratories to check at least the most critical false response rate for a qualitative test.

• It is reasonable to expect laboratories to be following published codes of best practice in qualitative testing where they are available.

• Quantitative (i.e. numerical) reports of uncertainties in qualitative test results should not generally be expected.