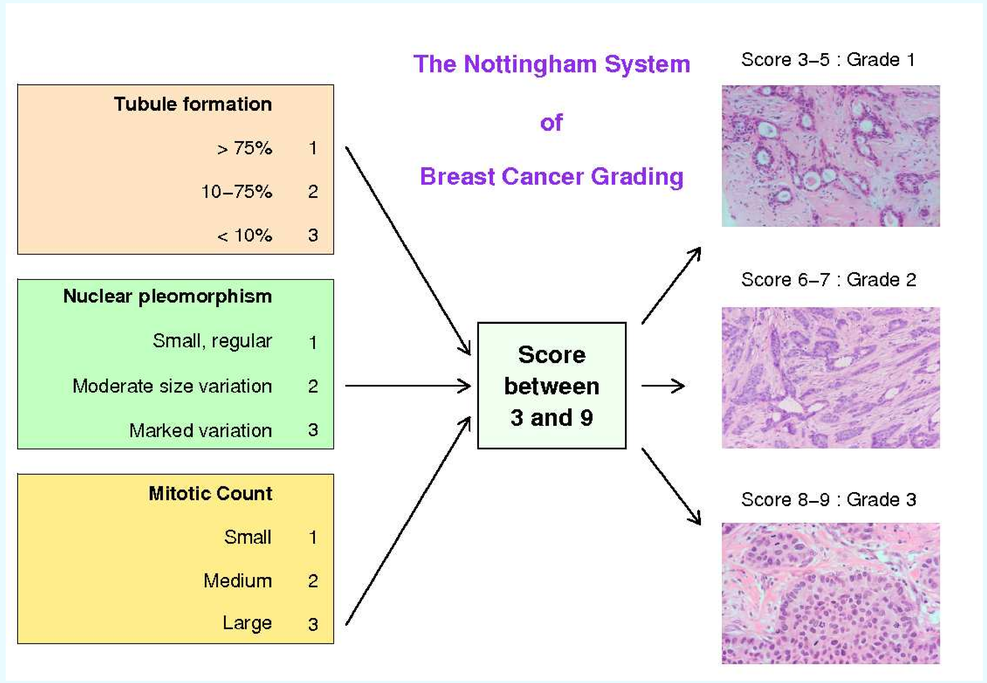
Assessing variability in histological grade of breast cancer tumours

T.R. Fanshawe, CHICAS, Department of Medicine, Lancaster University: t.fanshawe@lancaster.ac.uk

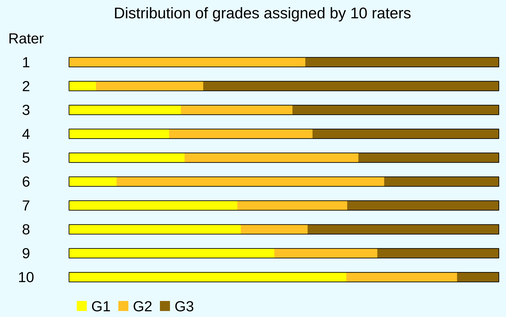


Data from the National External Quality Assurance (EQA) Scheme for Breast Cancer Histopathology:

- Digitised images of histological tissue sections of 52 breast cancer tumours
- 732 UK pathologists (raters) gave grades (G1, G2 or G3) to an average of 33 tumours each (63%)
- Each tumour was graded by between 390 (53%) and 513 (70%) raters
- No 'Ground Truth': grading is subjective

Are All Raters Equal?

- It seems not
- Big differences in distribution of grades awarded
- Rater 1 gave 0 Grade 1s out of 20 tumours
- Rater 10 gave 20 Grade 1s out of 31 tumours
- Different raters may have seen different tumours

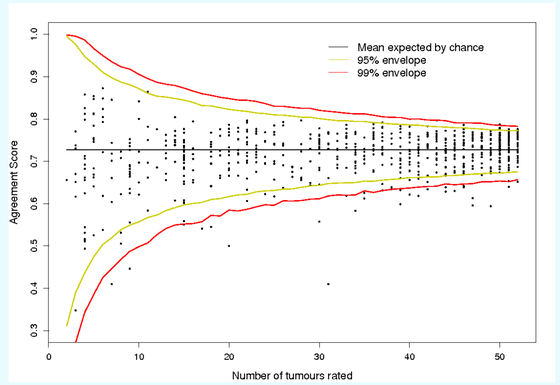


Acknowledgements

A.G. Lynch, I.O. Ellis, A.R. Green, R. Hanka
The EQA scheme is funded by the NHS Breast Screening Program
Fanshawe, Lynch *et al*: PLoS ONE 3(8): e2925. doi:10.1371/journal.pone.0002925

A Simple Summary Measure

- For each rater, calculate the proportion of raters with whom he agrees on a given tumour
- Average these across all tumours graded by the rater
- Plot values from all raters as a funnel plot (right)
- 95% and 99% envelopes estimated by Monte Carlo simulation
- Too many points outside envelopes: evidence that some raters are doing worse than others



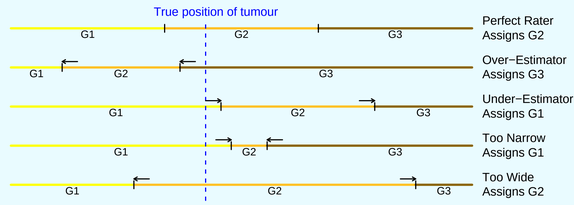
Cumulative Logit Model : The Idea

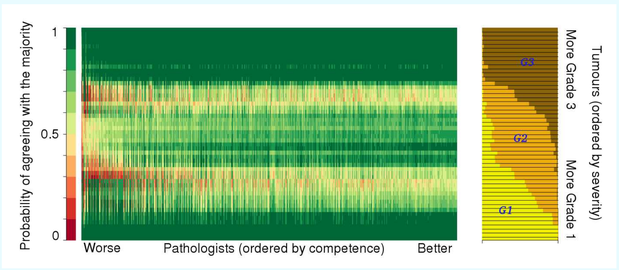
- Each tumour has a 'true severity' μ on a hypothetical continuous scale
- There are also 'true' boundaries b_{12} and b_{23} on the scale that dictate the 'true' grade
  i.e. $\mu < b_{12} \Rightarrow$ G1; $b_{12} < \mu < b_{23} \Rightarrow$ G2; $\mu > b_{23} \Rightarrow$ G3
- Each rater has his own set of boundary parameters that determine which grade he will assign (below right)

For rater j and tumour i:

$$\mathrm{P}(j \text{ assigns G3 to } i) = \mathrm{logit}^{-1}(f(i,j))$$
$$\mathrm{P}(j \text{ assigns G2 or G3 to } i) = \mathrm{logit}^{-1}(g(i,j))$$
$$f(i,j) = \lambda_i(\mu_i - b_{23,j})$$
$$g(i,j) = \lambda_i(\mu_i - b_{12,j})$$

(λ_i) are 'clarity' parameters, (μ_i) are 'severity' parameters
$(b_{12,j})$ and $(b_{23,j})$ are 'boundary' parameters





- Why do raters disagree?
- Graph shows estimated boundaries b_{12} and b_{23} for each rater
- Most common discrepancy is for raters to either under-estimate or over-estimate both boundaries

Results

- Sampling from posterior distributions enables quantities of interest to be estimated
- Heat-map shows the estimated probability, for each tumour, of a rater agreeing with the majority
- Right-hand panel shows the estimated marginal distribution of grades for each tumour