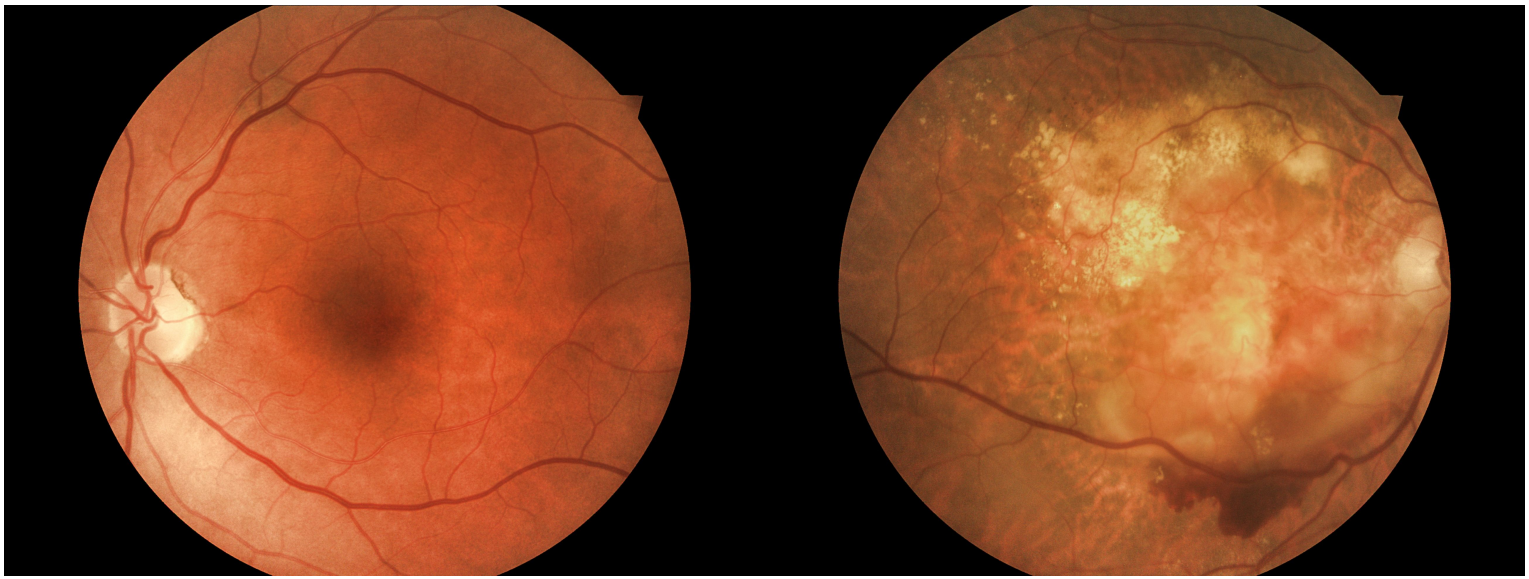
Measurement Error and Misclassification in statistical models: case studies bcam Bilbao

Helmut Küchenhoff
Statistical Consulting Unit
Ludwig-Maximilians-Universität München

Bilbao
31-05-2019

Age-related macular degeneration AMD

- AMD: Degenerative disorder of the macular and leading cause of irreversible blindness in elderly people of industrialised countries
- Accumulation of extracellular material (*drusen*) between specific retinal layers



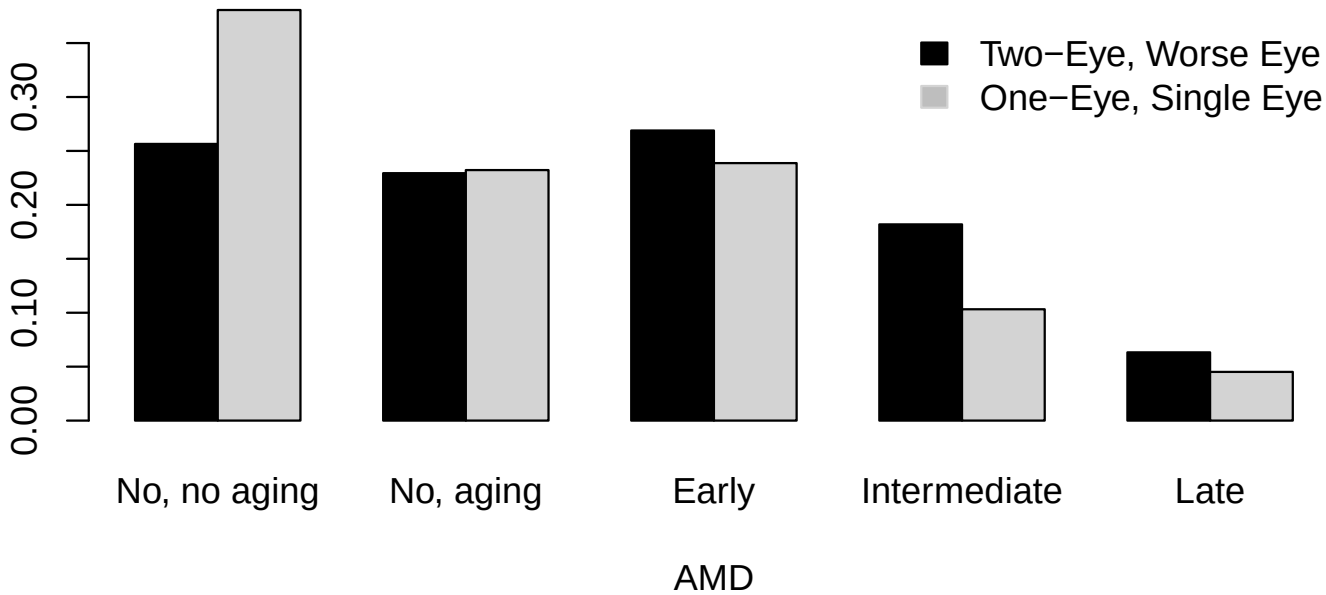
The German AugUR study

- Age related diseases: understanding genetic and non-genetic associations – a study at the Universität Regensburg (Stark et al. 2015).
- Prospective population based study in the (mobile) elderly population around Regensburg (Bavaria) to investigate age-related diseases
- Baseline survey 2013-2015 with 1133 participants
- 1040 participants with gradable fundus images in at least one eye
- Approx. 15% (148) participants have a missing single eye grading

Misclassification structure

- AMD grading is in general done per eye
- Standard in epidem. literature: AMD disease stage of a person defined by disease stage of worse eye
  Binary: $Y_i := max(Z_{1i}, Z_{2i}),\ Z_{1i}, Z_{2i} \in \{0 = \text{no}, 1 = \text{AMD}\}$
- $Y_i :=$ „Occurrence of AMD in at least one eye“
- Misclassification can occur because of (I) missing single eye gradings, (II) error in single eye gradings

AMD ocuurence depending on garding one or two eyes



Prevalence Estimation of AMD disease stages in Bavaria

- Sample proportions of age-sex groups differ from population proportions in Bavarian elderly
- To adjust for differential non-response: Estimation of AMD prevalence in each age-sex group and post stratification to Bavarian population
- Correction for disease stage misclassification in *one-eye participants* necessary
- Recall: with known (mis)classification probabilities true response class probabilities can be estimated based on observed misclassified response data:

$$\hat{\mathbb{P}}_Y = \Pi^{-1} \times \hat{\mathbb{P}}_{Y^*},$$

  where $\hat{\mathbb{P}}_{Y^*}$ can be estimated by observed class proportions („matrix method“)
- $\hat{\Pi}$ can also be estimated from validation data, matrix method can be derived as MLE of true class proportions if validation data is external, misclassification probabilities are constant and *transferable*

Other approach: Use positive or negative predicted values

Basic idea : Find model for

$$P(Y = 1|Y^* = i)$$

Then

$$P(Y = 1) = P(Y^* = 1)\cdot P(Y = 1|Y^* = 0) + P(Y^* = 0)P(Y = 1|Y^* = 0)$$

Calibration.

- Special structure in our case: $P(Y = 0|Y^* = 1) = 0$
- General case straight forward
- Transferability from validation data problematic

Prevalence estimation in AUGUR study

- If randomly selected internal validation sample of size n^v is available MLE of true class probabilities is given by

$$\hat{\mathbb{P}}_Y = \frac{n^v}{n}\hat{\mathbb{P}}^v_Y + \frac{n^m}{n}\hat{\Lambda} \times \hat{\mathbb{P}}^m_{Y^*},$$

  with
  - $\hat{\mathbb{P}}^v_Y$ ($\hat{\mathbb{P}}^m_{Y^*}$) MLE of multinomial class probabilities of the true (misclassified) response estimated based on the validation (main study) data
  - $n^m = n - n^v$ is the number of observations in the main study sample
  - $\hat{\Lambda} = [\hat{\lambda}_{kl}]$ is the matrix of *predictive values* $\lambda_{kl} = \mathbb{P}(Y = k|Y^* = l)$ estimated based on validation data
- Estimation of $\hat{\lambda}_{kl}$ by calculating proportion of all single eyes in disease stage l, for which worse eye is in disease stage k
- *Transition probability from random eye to worse eye disease stage*
- Assumption: Λ constant for all age-sex groups

Results

$$\hat{\Lambda}_{clin} = \begin{pmatrix} 0.718 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.172 & 0.790 & 0.000 & 0.000 & 0.000 \\ 0.084 & 0.168 & 0.861 & 0.000 & 0.000 \\ 0.025 & 0.040 & 0.124 & 0.923 & 0.000 \\ 0.000 & 0.003 & 0.014 & 0.077 & 1.000 \end{pmatrix}$$

AMD	$\hat{p}_k^{predval}$	$\hat{se}^{boot}$	$\hat{p}_k^{naive}$	$\hat{se}$
No, no aging	0.243	0.014	0.258	0.014
No, norm aging	0.231	0.015	0.228	0.014
Early	0.279	0.016	0.275	0.015
Intermediate	0.174	0.013	0.167	0.013
Late	0.074	0.010	0.072	0.010

Summary

Misclassification in Studies on AMD

- Naive modelling strategy ignoring missing disease stage gradings leads to biased estimates
- The bigger the fraction of missing observations, the bigger the bias

Uncertainty of psychiatric diagnosis

Mokros, A., Habermeyer, E.,Küchenhoff, H. (2018). The uncertainty of psychological and psychiatric diagnoses. Psychological Assessment.

- Psychiatric diagnosis are highly relevant
- this is especially true for the diagnosis of pedophilia
- estimation of positive predicted value and prevalence are relevant

Handling misclassification

- No gold standard is available
- κ is known from validation studies
- problems of identification

p :	Prevalence: probability of the occurrence of a disease (positive case)
p^* :	Observed prevalence: probability of rating a case as positive
sens :	Sensitivity: probability of identifying a positive case correctly
spec :	Specificity: probability of identifying a negative case correctly
ppv :	Positive predictive value: probability that case rated as positive is truly positive
κ	Cohen's kappa: Chance–corrected inter–rater reliability of two raters

In general, the following equations hold:

$$p^* = p \cdot sens + (1 - p)(1 - spec) \quad (1)$$

$$ppv = \frac{p \cdot sens}{p \cdot sens + (1 - p)(1 - spec)} = \frac{p \cdot sens}{p^*} \quad (2)$$

identification regions

Assuming that two raters score independently with identical *sens* and *spec*, the following equation holds for κ

$$\kappa = \frac{p\,(1-p)\,(sens+spec-1)^2}{(spec-p\,(sens+spec-1))} \cdot \frac{1}{(1-spec+p\,(sens+spec-1))} \tag{3}$$

For a given observed prevalence p^* and κ, Küchenhoff et al.(2012) deduce identification regions for the true prevalence p, sensitivity and specificity:

$$I(p \parallel p^*, \kappa) = \left[\frac{p^*}{p^* + \kappa^{-1}(1-p^*)};\quad \frac{p^*}{p^* + \kappa\,(1-p^*)} \right], \tag{4}$$

$$I(sens \parallel p^*, \kappa) = [p^* + \kappa\,(1-p^*);\, 1] \tag{5}$$

$$I(spec \parallel p^*, \kappa) = [1-p^* + p^*\kappa;\, 1]\,. \tag{6}$$

Identification region for *ppv*

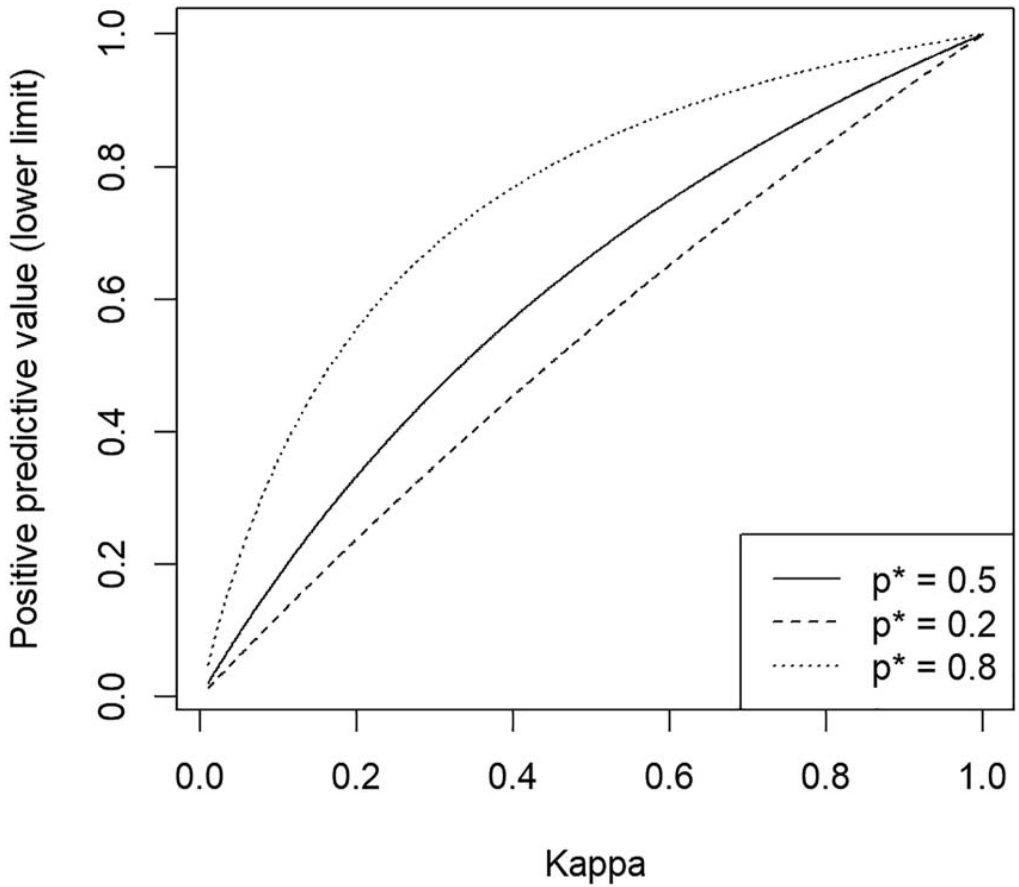
Note that the lower limit of the prevalence identification region corresponds to the case of $sens = 1$ and the upper limit corresponds to $spec = 1$.

$$ppv_l = \frac{p_l \cdot sens_u}{p^*} = \frac{p^* \left[p^* + \kappa^{-1}(1-p^*)\right]^{-1}}{p^*} = \left[p^* + \kappa^{-1}(1-p^*)\right]^{-1} \tag{7}$$

The upper limit of the identification region of *ppv* is 1, since it corresponds to the case of $spec = 1$. The estimated identification region is

$$I(spec \parallel \hat{p}^*, \hat{\kappa}) = \left[\left[\hat{p}^* + \hat{\kappa}^{-1}(1-\hat{p}^*)\right]^{-1}; 1\right] \tag{8}$$

Relationship between κ ans ppv



Results

Naive prevalence estimate:	$\hat{p}^* = 0.4 \ \ (SE = 0.017)$
Estimate of kappa:	$\hat{\kappa} = 0.65 \ \ (SE = 0.049)$
Lower bound of sensitivity :	0.79
Lower bound of specificity :	0.86
Interval for prevalence:	$[0.30, 0.51]$
95% CI of the true prevalence:	$[0.27, 0.54]$
Interval for PPV:	$[0.76, 1]$
The 95% CI of the PPV:	$[0.66, 1]$

Conclusion

- Misclassification has a relevant impact
- Reasoning about uncertainty of measurements