

1. INTRODUCTION

Abstract. The concordance correlation coefficient (Lin, 1989) has been defined as a specific intraclass correlation coefficient when subjects are considered as a random effect and observers as a fixed effect (Carrasco and Jover, 2003). Using this result, the concordance correlation coefficient was extended for measuring agreement between counts by means of the intraclass correlation coefficient derived from the Poisson-Normal generalized linear mixed model (Carrasco and Jover, 2005), where the link function is the logarithm, the variability between subjects is Normal distributed and the within-subject variation follows a Poisson distribution.

However, it would be possible to find overdispersion when analyzing this kind of data. In this case, the assumption that within-subject variability is Poisson distributed would not hold, and the expression of the concordance correlation coefficient should be accordingly modified. In this work we show two alternative expressions of the concordance correlation coefficient for count data to account for overdispersion. In the first solution the within-subject variance is expressed as a product of the scale parameter and the mean, whereas in the second the assumption about the within-subject distribution is changed from Poisson to Negative Binomial. The three concordance correlation coefficient expressions are estimated and compared in a data example set where the counts of CD34+ cells are obtained using different techniques.

Data. Two or more observers rate n subjects in a discrete scale. The outcome consist is counts.

Index of between-observers agreement. The CCC is defined as the specific Intraclass Correlation Coefficient as the ratio of the covariance between data from same subject and different observer to the overall variance of data

$$\rho = \frac{\text{cov}(Y_{ij}, Y_{ji})}{\text{Var}(Y_{ij})}$$

The cluster is the subject and the overall variance must take into account all the within-subject sources of variability: between-observers and random error.

Models considered. Y_{ij} : outcome taken by the j th observer to the i th subject; β_0 : intercept; α_i : subject effect; β_j : observer effect; $i=1, \dots, n$; $j=1, \dots, k$

1) Poisson- Normal Mixed Model

$$Y_{ij} | \alpha_i, \beta_j \sim \text{Poisson}(\mu_{ij}) \quad \alpha_i \sim N(0, \sigma_\alpha) \quad \log(\mu_{ij}) = \beta_0 + \alpha_i + \beta_j \quad E(Y_{ij} | \alpha_i, \beta_j) = \mu_{ij} \quad \text{Var}(Y_{ij} | \alpha_i, \beta_j) = \mu_{ij}$$

$$E(Y_{ij}) = \exp\left(\beta_0 + \frac{\sigma_\alpha^2 + \sigma_\beta^2}{2}\right) \quad \text{Var}(Y_{ij}) = E[Y_{ij}] \cdot \left\{ E[Y_{ij}] \left(e^{\sigma_\alpha^2 + \sigma_\beta^2} - 1 \right) + 1 \right\} \quad \sigma_\beta^2 = \frac{1}{k} \sum_{j=1}^k \beta_j^2 \quad \text{cov}(Y_{ij}, Y_{ji}) = E[Y_{ij}]^2 \cdot (e^{\sigma_\alpha^2} - 1)$$

$$\rho = \frac{E[Y_{ij}] \cdot (e^{\sigma_\alpha^2} - 1)}{E[Y_{ij}] \cdot (e^{\sigma_\alpha^2 + \sigma_\beta^2} - 1) + 1}$$

2) Overdispersed Poisson- Normal Mixed Model (ϕ = scale parameter)

$$Y_{ij} | \alpha_i, \beta_j \sim \text{Poisson}(\mu_{ij}) \quad \alpha_i \sim N(0, \sigma_\alpha) \quad \log(\mu_{ij}) = \beta_0 + \alpha_i + \beta_j \quad E(Y_{ij} | \alpha_i, \beta_j) = \mu_{ij} \quad \text{Var}(Y_{ij} | \alpha_i, \beta_j) = \phi \cdot \mu_{ij}$$

$$E(Y_{ij}) = \exp\left(\beta_0 + \frac{\sigma_\alpha^2 + \sigma_\beta^2}{2}\right) \quad \text{Var}(Y_{ij}) = E[Y_{ij}] \cdot \left\{ E[Y_{ij}] \left(e^{\sigma_\alpha^2 + \sigma_\beta^2} - 1 \right) + \phi \right\} \quad \sigma_\beta^2 = \frac{1}{k} \sum_{j=1}^k \beta_j^2 \quad \text{cov}(Y_{ij}, Y_{ji}) = E[Y_{ij}]^2 \cdot (e^{\sigma_\alpha^2} - 1)$$

$$\rho = \frac{E[Y_{ij}] \cdot (e^{\sigma_\alpha^2} - 1)}{E[Y_{ij}] \cdot (e^{\sigma_\alpha^2 + \sigma_\beta^2} - 1) + \phi}$$

3) Negative Binomial- Normal Mixed Model (k = dispersion parameter)

$$Y_{ij} | \alpha_i, \beta_j \sim \text{NegBin}(\mu_{ij}, k) \quad \alpha_i \sim N(0, \sigma_\alpha) \quad \log(\mu_{ij}) = \beta_0 + \alpha_i + \beta_j \quad E(Y_{ij} | \alpha_i, \beta_j) = \mu_{ij} \quad \text{Var}(Y_{ij} | \alpha_i, \beta_j) = \mu_{ij} + k \cdot \mu_{ij}^2$$

$$E(Y_{ij}) = \exp\left(\beta_0 + \frac{\sigma_\alpha^2 + \sigma_\beta^2}{2}\right) \quad \text{Var}(Y_{ij}) = E[Y_{ij}] \cdot \left\{ E[Y_{ij}] \left[(k+1) \cdot e^{\sigma_\alpha^2 + \sigma_\beta^2} - 1 \right] + 1 \right\} \quad \sigma_\beta^2 = \frac{1}{k} \sum_{j=1}^k \beta_j^2 \quad \text{cov}(Y_{ij}, Y_{ji}) = E[Y_{ij}]^2 \cdot (e^{\sigma_\alpha^2} - 1)$$

$$\rho = \frac{E[Y_{ij}] \cdot (e^{\sigma_\alpha^2} - 1)}{E[Y_{ij}] \cdot \left[(k+1) \cdot e^{\sigma_\alpha^2 + \sigma_\beta^2} - 1 \right] + 1}$$

Estimation Approach.

Gaussian Quadrature using NLMIXED SAS Procedure for Models 1 and 3.

Penalized Quasi-Likelihood using GLIMMIX SAS Procedure for Model 2.

Confidence intervals are built using asymptotic normal theory. Standard errors are derived through the delta method

3. CASE EXAMPLE AND RESULTS

Outcome: CD34+ cells released from apheresis products collected from 20 subjects using two methods: SYTO-13 and Procount (Fornas *et al.* 2000).

Aim: To measure the agreement between the two CD34+ counting methods.

From **Table 1** the point estimates are quite similar but standard errors from Overdispersed Poisson and Negative Binomial models are greater giving wider confidence intervals.

To analyze the implication of ignoring the overdispersion a simulation study of 2000 runs was carried out. It was considered the Negative Binomial model as the true model using their estimates as parameters, thus the true CCC was 0.832. The sample size was set at 30 and 100 subjects. **Table 2** shows the simulation results.

	Model		
	Poisson-Normal	Overdispersed Poisson-Normal	Negative Binomial - Normal
β_0	6.448 (0.249) [5.928 : 6.969]	6.430 -----	6.430 (0.249) [5.908 : 6.951]
β_1	-0.489 (0.015) [-0.514 : -0.466]	-----	-0.409 (0.073) [-0.562 : -0.256]
σ_α^2	1.235 (0.392) [0.415 : 2.055]	-----	1.189 (0.385) [0.382 : 1.996]
k	-----	-----	0.049 (0.016) [0.015 : 0.083]
ρ	0.847 (0.022) [0.802 : 0.893]	0.829 (0.035) [0.760 : 0.894]	0.832 (0.045) [0.737 : 0.927]
Deviance	548.2	-----	20.66
Scale	28.85	-----	1.09
AIC	1038.4	-----	576.9

	Model					
	Poisson-Normal		Overdispersed Poisson-Normal		Negative Binomial - Normal	
	$n = 30$	$n = 100$	$n = 30$	$n = 100$	$n = 30$	$n = 100$
Mean of Estimates	0.879	0.887	0.852	0.858	0.823	0.830
Bias	0.048	0.055	0.020	0.026	-0.008	-0.002
Relative Bias (%)	5.75	6.61	2.44	3.19	-1.00	-0.21
Mean Squared Error (%)	0.487	0.374	0.348	0.156	0.169	0.041
Mean of Estimated Variances (x1000)	0.279	0.065	0.876	0.253	1.526	0.418
Variance of Estimates (x1000)	2.583	0.722	3.071	0.854	1.627	0.410
Coverage (%)	34.37	10.34	61.18	51.28	94.45	95.35

4. CONCLUSIONS AND MAIN REMARKS

- 1) The overdispersion can be accommodated by means of an overdispersed Poisson model where the variance is function of the mean and the scale parameter, or changing the assumption of the within-subject probability distribution to Negative Binomial.
- 2) In the case example the Poisson-Normal model gave a great overdispersion with a scale parameter far from 1 whereas the model using the Negative-Binomial behaves very much better.
- 3) The main consequence of ignoring overdispersion is a poor estimate of the standard error of the CCC.
- 4) Although the Overdispersed Poisson Model incorporates the overdispersion in the point estimate of the CCC, the standard error is still underestimated.
- 5) The estimates from the Poisson and Overdispersed Poisson seem to be biased because the bias do not reduce as sample size increases.
- 6) The underestimation of the standard error leads to a poor coverage. Even though the confidence intervals are approximately well centered, they are narrower than desired.
- 7) Thus, to measure agreement between observers when the outcome consists in counts using the Poisson-Normal mixed model, one would have to be reasonably sure that the overdispersion is small, otherwise the confidence interval might be misleading. In that case, the Negative Binomial is a good alternative to the Poisson-Normal model.

5. REFERENCES

- Lin L. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; **45**: 255-268
- Fornas O, Garcia J, Petriz J. (2000). Flow cytometry counting of CD34+ cells in whole blood. *Nature medicine*, **6**, 833-836.
- Carrasco JL, Jover L. (2003). Estimating the generalized concordance correlation coefficient through variance components. *Biometrics*, **59**, 849-858.
- Carrasco JL, Jover L. (2005). Concordance correlation coefficient applied to discrete data. *Statistics in Medicine*, **24**: 4021-4034