COMPARISON OF CONVENTIONAL REGRESSION AND GENERALIZED ESTIMATING EQUATIONS (GEE) FOR THE ANALYSIS OF LONGITUDINAL DATA: AN APPLICATION TO WORK-RELATED RISK FACTORS FOR LOW BACK PAIN

Hoogendoorn, W.E.[1,2,3], Bongers, P.M[1], De Vet, H.C.W.[2], Twisk, J.W.R.[2], Van Mechelen, W.[2,3], Bouter, L.M.[2]

1 TNO Work and Employment, Hoofddorp, The Netherlands
2 Institute for Research in Extramural Medicine, Faculty of Medicine, Vrije Universiteit, Amsterdam, The Netherlands
3 Department of Social Medicine, Faculty of Medicine, Vrije Universiteit, Amsterdam, The Netherlands

Aim
To compare the results of conventional regression and generalized estimating equations (GEE) analysis in a three-year prospective cohort study on work-related risk factors for low back pain.

Methods
The study population consisted of a cohort of 1,192 workers with no low back pain at baseline. Both at baseline and at the three annual follow-up measurements, information on work-related physical and psychosocial factors and the occurrence of low back pain was obtained by means of questionnaires. In a conventional logistic regression model, physical and psychosocial risk factors at baseline were related to the cumulative incidence of low back pain during the three-year follow-up period. In a GEE logistic model, repeated measurements of the physical and psychosocial risk factors were related to low back pain reported at one measurement point later.

Results
The conventional regression model showed a statistically significant effect of flexion and/or rotation of the upper part of the body (OR=1.8; 95% CI=1.2-3.0), but not of moving heavy loads (OR=1.4; 95% CI=0.7-3.1). The GEE model showed a somewhat greater and statistically significant effect of both flexion and/or rotation of the upper part of the body (OR=2.2; 95% CI=1.5-3.2) and moving heavy loads (OR=1.6; 95% CI=1.0-2.6). With both methods no statistically significant associations with low back pain were found for the psychosocial work characteristics, but the GEE model showed weaker odds ratios for these variables than the conventional regression model.

Conclusions
The results show that there are differences between the two analytical methods in both the magnitude and the precision of the observed odds ratios. The interpretation of these differences will also be the subject of this presentation.