Sequential Testing of Measurement Errors in Inter-Rater Reliability Studies

Mei Jin\textsuperscript{1}, Aiyi Liu\textsuperscript{2}, Zhen Chen\textsuperscript{2}, and Zhaohai Li\textsuperscript{1}

\textit{Department of Statistics, George Washington University, Washington, DC, U.S.A.}\textsuperscript{1}

\textit{National Institute of Child Health and Human Development, Bethesda, MD, U.S.A.}\textsuperscript{2}

\textit{Abstract:} Inter-rater reliability is usually assessed by means of the intraclass correlation coefficient. Using two-way analysis of variance to model raters and subjects as random effects, we derive group sequential testing procedures for the design and analysis of reliability studies in which multiple raters evaluate multiple subjects. Compared with the conventional fixed sample procedures, the group sequential test has smaller average sample number. The performance of the proposed technique is examined using simulation studies and critical values are tabulated for a range of two-stage design parameters. The methods are exemplified using data from the Physician Reliability Study for diagnosis of endometriosis.

\textit{Key words and phrases:} Interim analysis; intraclass correlation coefficient; inter-rater reliability; measurement errors; sample size and power; two-way ANOVA.