Letter to the Editor

Factors that Impact on Interrater Reliability of the Mouse Clinical Frailty Index

Alice E. Kane,1 Sarah N. Hilmer,1 Aniko Huizer-Pajkos,2 John Mach,3 Dawn Nines,4 Dawn Boyer,5 Kristan Gavin,4 Sarah J. Mitchell,7 and Rafael de Cabo6

1Department of Clinical Pharmacology and Aged Care, University of Sydney, Royal North Shore Hospital, Sydney, New South Wales, Australia. 2Department of Aged Care, Kolling Institute of Medical Research, Royal North Shore Hospital, Sydney, New South Wales, Australia. 3Laboratory of Ageing and Pharmacology, Kolling Institute of Medical Research, Royal North Shore Hospital, Sydney, New South Wales, Australia. 4Translational Gerontology Branch, National Institute on Aging, Baltimore, Maryland. 5Biomedical Research Center, National Institute on Aging, NIH, Baltimore, Maryland. 6Experimental Gerontology Section, National Institute on Aging, Baltimore, Maryland. 7Laboratory of Experimental Gerontology, National Institute on Aging, Baltimore, Maryland.

Address correspondence to Alice E. Kane, Department of Clinical Pharmacology and Aged Care, University of Sydney, Royal North Shore Hospital, Sydney, New South Wales, Australia. Email: alice.kane@sydney.edu.au

Dear Editor,

We read with great interest the article “Reliability of a Frailty Index Based on the Clinical Assessment of Health Deficits in Male C57BL/6 Mice” recently published online in the Journal of Gerontology Biological Sciences (1). Feridooni and colleagues investigated the reliability of a recently developed mouse clinical frailty index between two raters, in three separate groups of male C57BL/6 mice aged 0.9–1.2 years, with discussion and refinement of index descriptors between assessment of each group. They found high overall inter-rater reliability, and that measures of reliability increased sequentially over the three groups. Our assessment of the mouse clinical frailty index in one published (2) and two unpublished studies of male C57BL/6 mice shed further light on factors that impact on its inter-rater reliability.

Our studies clarify that practice and experience with assessment of the mouse clinical frailty index, without discussion or refinement, does not improve inter-rater reliability. We have completed two projects assessing the frailty index of old (Project 1 n = 55, 26.1 ± 0.5 months; Project 2 n = 17, 23.9 ± 0.2 months) mice before and after 4–6 weeks of a dietary/pharmaceutical intervention. Each project used two raters for the frailty index, who did not discuss or compare their scoring until the completion of the projects. The inter-rater reliability between the two raters for both of these projects did not improve with practice or experience (Project 1 Pre ICC=0.522 [CI 0.181–0.721], p = .00; r² = .361, p = .00; Post ICC = 0.488 [CI 0.244–0.652], p = .01; r² = .328, p = .00; Project 2 Pre ICC = 0.705 [CI 0.186–0.893], p = .01; r² = .562, p = .02; Post ICC = 0.561 [CI 0.122–0.781], p = .01; r² = .391, p = .02).

These results fall within the range of reliability statistics observed by Feridooni and colleagues (1), using different raters in different research groups and countries.

Another factor that we have identified as affecting the inter-rater reliability when assessing the mouse clinical frailty index is the professional background and baseline animal training of the raters. In order to explore this, we had four raters assess the clinical frailty index of a cohort of old (n = 74, 19.0 ± 1.0 months) mice, with no discussion or training period. Two raters were animal technicians with 5–10 years of experience and training in animal handling, and two raters were research scientists with minimal animal handling training, and 3–6 years of animal research experience. Interestingly, the inter-rater reliability between the technicians and researchers was poor (ICC = 0.201–0.489, p < .05; r² = –0.028 to 0.33, p < .05), despite moderate inter-rater reliability between the technicians (ICC = 0.605 [CI 0.344–0.762], p = .00; r² = .454, p = .00) and excellent inter-rater reliability between the scientists (ICC=0.88 [95%CI 0.80–0.92] p = 0.00; r² = .817, p = .00). Furthermore, it does appear to be the background, rather than the years of experience that are important, as the inter-rater reliabilities in our pre and post intervention studies (data above), which were all performed by scientists, were moderate to high despite the raters including PhD students and a senior postdoc/laboratory manager. This suggests that the selection of raters for the mouse clinical frailty index must be done carefully, and preference given to raters with the same training/background.

© The Author 2015. Published by Oxford University Press on behalf of The Gerontological Society of America. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
but not necessarily experience level, if comparisons are to be made between and across mouse groups and studies.

References