Postal and Face-to-Face Administration of Stroke Outcome Measures
Can Mixed Modes Be Used?

Christopher J. Sutton, PhD; Caroline L. Watkins, PhD; Neil Cook, MSc; Michael J. Leathley, PhD; Joanna McAdam, MA; Paola Dey, MD

Background and Purpose—Different modes of administration are used to collect stroke outcomes, even within the same study, potentially leading to different results. We investigated the effect of administration mode (postal questionnaire; face-to-face interview) on self-reports of activities of daily living and mood.

Methods—The study was nested within a poststroke motivational interviewing trial. Activities of daily living (Barthel; Nottingham Extended) and mood (General Health Questionnaire; Yale) were collected at 3 and 12 months via postal questionnaire. Participants were approached to respond again via face-to-face interview. Paired t tests (McNemar test) and intraclass correlation coefficients (Cohen \( \kappa \)) were used, with 95% CI, to compare scores (items).

Results—Forty-four participants consented. Only Barthel scores were significantly different; they were 1.0 (95% CI, 0.5–1.6) higher face-to-face. The intraclass correlation coefficient for the Barthel was 0.90; for the other scales it was between 0.83 and 0.87. The Yale \( \kappa \) was 0.72.

Conclusions—Modes of administration might be used interchangeably, albeit in conjunction with corrections for the Barthel.

Key Words: activities of daily living ■ mood ■ outcomes ■ reproducibility

Stroke recovery is affected by limitations in activities of daily living (ADL) and mood,1 and stroke treatment and care evaluations include measures of function and emotion. Validated questionnaires are used because collection of directly observed performance measures is often impractical and direct observation of mood has limited success.2 However, delivery mode can affect their measurement properties. Question meanings can be clarified in face-to-face interviews, but these are expensive and interviewers may misinterpret responses.3 Postal administration is cheaper and perceived anonymity may enhance sensitive information disclosure, but questionnaire completion may be compromised.3 Stroke adds further complexity: patient interpretation or communication is often compromised and upper limb hemiplegia makes self-completion difficult. We found no studies comparing postal and interviewer administration for poststroke mood outcomes and few for other constructs.4–6 This study compares ADL and mood responses from postal questionnaire and face-to-face interview, and explores factors which may explain response differences.

Methods
The study was conducted within a poststroke motivational interviewing randomized controlled trial.7 Ethical approval was obtained. Baseline measures, collected face-to-face by a research nurse, included cognition (Rivermead Behavioural Memory Test),8 communication (Frenchay Aphasia Screening Test),9 mood (12- and 28-item General Health Questionnaire9; Yale),9 and function (Barthel Index).10 Outcomes collected 3- and 12-months poststroke by postal questionnaire included General Health Questionnaire-12, General Health Questionnaire-28, Barthel Index,10 Nottingham Extended ADL11 and Yale.

Trial participants returning 3- or 12-month postal questionnaires between January 2005 and February 2006 were flagged, without disclosing their data, by a research assistant to an experienced research nurse. Those recruited at 3 months were not approached again at 12 months. The nurse arranged a timely face-to-face interview (within 14 [31] days of questionnaire receipt for 3-[12]-month follow-up) and read out questions identical to the postal questionnaire, recording participants’ responses.

Sample size target was 50 so that 95% CIs for intraclass correlation coefficients of scale scores would have width of <0.2, assuming intraclass correlation coefficient is 0.8. Differential bias was assessed using 95% CI based on paired t tests (McNemar test for Yale). Repeatability was measured using intraclass correlation coefficient for scales and Cohen \( \kappa \) for items (with quadratic weighting when ordinal), with 95% CI, using bias-corrected bootstrapping for \( \kappa \). Scale scores were compared using Bland-Altman and scatter plots. For scales, if repeatability was inadequate we performed item analysis, and if differential bias was detected we used exploratory linear modeling to investigate whether participant characteristics could explain between-mode differences.

Results
Postal questionnaires were returned by 109 trial participants during the recruitment period, of whom 44 (40.4%) were interviewed (Table 1).
Only the Barthel showed a significant between-mode difference, with mean scores 1.0 (95% CI, 0.5–1.6) higher face-to-face (Table 2); differences appeared larger among participants with lower scores (Figure). Repeatability was generally high (Table 2).

Table 2. Differences and Repeatability for Outcomes and Barthel Items (n=44)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean (SD) Postal Score</th>
<th>Mean (SD) Face-to-Face Score</th>
<th>Mean Difference (95% CI)</th>
<th>Repeatability (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GHQ-28 (n=43)*</td>
<td>7.6 (8.0)</td>
<td>7.5 (7.5)</td>
<td>0.1 (−1.3 to 1.5)</td>
<td>0.83† (0.71 to 0.90)</td>
</tr>
<tr>
<td>GHQ-12 (n=43)*</td>
<td>4.5 (4.5)</td>
<td>4.3 (4.1)</td>
<td>0.1 (−0.6 to 0.8)</td>
<td>0.86‡ (0.75 to 0.92)</td>
</tr>
<tr>
<td>Yale: often feeling sad/depressed (n=43)*</td>
<td>22 (51.2%)</td>
<td>20 (46.5%)</td>
<td>4.7% (−8.8% to 18.1%)</td>
<td>0.72‡ (0.50 to 0.91)</td>
</tr>
<tr>
<td>NEADL</td>
<td>12.6 (6.5)</td>
<td>13.0 (5.4)</td>
<td>−0.4 (−1.3 to 0.5)</td>
<td>0.87† (0.78 to 0.93)</td>
</tr>
<tr>
<td>Barthel</td>
<td>16.5 (4.4)</td>
<td>17.5 (3.5)</td>
<td>−1.0 (−1.6 to −0.5)</td>
<td>0.90† (0.83 to 0.95)</td>
</tr>
<tr>
<td>Dressing</td>
<td>1.64 (0.65)</td>
<td>1.68 (0.64)</td>
<td>−0.05 (−0.18 to 0.08)</td>
<td>0.78‡ (0.53 to 0.92)</td>
</tr>
<tr>
<td>Toilet</td>
<td>1.82 (0.50)</td>
<td>1.86 (0.46)</td>
<td>−0.05 (−0.11 to 0.02)</td>
<td>0.90‡ (0.40 to 1)</td>
</tr>
<tr>
<td>Bathing</td>
<td>0.52 (0.51)</td>
<td>0.59 (0.50)</td>
<td>−0.07 (−0.15 to 0.01)</td>
<td>0.86‡ (0.69 to 1)</td>
</tr>
<tr>
<td>Transfers</td>
<td>2.70 (0.59)</td>
<td>2.80 (0.67)</td>
<td>−0.09 (−0.22 to 0.04)</td>
<td>0.77‡ (0.43 to 0.88)</td>
</tr>
<tr>
<td>Mobility</td>
<td>2.73 (0.69)</td>
<td>2.84 (0.64)</td>
<td>−0.11 (−0.23 to 0.00)</td>
<td>0.66‡ (−0.04 to 0.91)</td>
</tr>
<tr>
<td>Grooming</td>
<td>0.84 (0.37)</td>
<td>0.98 (0.15)</td>
<td>−0.14 (−0.24 to −0.03)</td>
<td>0.22‡ (0.00 to 0.73)</td>
</tr>
<tr>
<td>Urinary function</td>
<td>1.32 (0.77)</td>
<td>1.50 (0.73)</td>
<td>−0.18 (−0.36 to 0.00)</td>
<td>0.68‡ (0.43 to 0.88)</td>
</tr>
<tr>
<td>Bowel function</td>
<td>1.70 (0.59)</td>
<td>1.89 (0.32)</td>
<td>−0.18 (−0.36 to 0.00)</td>
<td>0.24 (0.02 to 0.64)</td>
</tr>
<tr>
<td>Feeding</td>
<td>1.75 (0.49)</td>
<td>1.93 (0.26)</td>
<td>−0.18 (−0.33 to −0.03)</td>
<td>0.17‡ (−0.04 to 0.55)</td>
</tr>
</tbody>
</table>

GHQ indicates General Health Questionnaire; NEADL, Nottingham Extended Activities Of Daily Living.

*Number with data available.
†Intraclass correlation coefficients.
‡κ.

Discussion

This study suggests consistency between stroke patient mood and ADL questionnaire responses postally and face-to-face. However, Barthel scores were often higher face-to-face, resulting in a 1-point significant difference in means. People may be more likely to self-report disability postally, suspecting it might affect benefit entitlement; social desirability bias might affect face-to-face interviews.3 Potentially embarrassing questions around incontinence showed relatively large biasing effects, albeit of borderline significance. Some differences between delivery modes have previously been found in item responses on other ADL questionnaires.4 Seven postal questionnaires were completed with support from carers; proxies have been shown to rate outcome worse than stroke patients.12

Barthel delivery mode could affect conclusions of disability surveys, but in randomized controlled trials its influence would not differ between arms, unless assessment bias was different; there was no intervention group effect on between-mode differences in our study. Alternatively, assessment bias may result in ceiling effects, a recognized problem with the Barthel,13 potentially reducing treatment effects; we found smaller between-mode differences in participants near its ceiling. If face-to-face
interviewing is used to supplement postal collection, it would be prudent to consider potential biases.

Our study is not without limitations. Postal questionnaires were always completed first, but most participants were 12-months poststroke and follow-up time was not related to differential bias, so functional improvement is unlikely to explain our findings. Most participants had high Barthel scores and, because of the trial intervention, no patient had severe cognitive or communication problems. Although our sample was small and homogeneous, we detected factors that might help explain Barthel differences. We were unable to conclude which mode is more valid as there was no objective ADL assessment.

We therefore recommend further research, using larger, more comprehensive samples, and randomized design, into using different delivery modes for the Barthel, including telephone interviewing not investigated here. Research should include investigating corrections for potential differential bias introduced via mixed assessment methods. In the interim, care should be taken in mixing modes of Barthel administration, particularly when differential bias cannot be reasonably assumed to cancel out.

Acknowledgments
The authors acknowledge Helen Gardner and Irene McClelland (data collection), Denise Forshaw (data coordination/input), and Jenny Marsden (data management).

Sources of Funding
This study was funded by National Health Service (NHS) Executive, North West; Royal Liverpool and Broadgreen University Hospitals NHS Trust; and University of Central Lancashire.

Figure. Comparison of Barthel scores face-to-face and postally.

Disclosures
None.

References
Postal and Face-to-Face Administration of Stroke Outcome Measures: Can Mixed Modes Be Used?
Christopher J. Sutton, Caroline L. Watkins, Neil Cook, Michael J. Leathley, Joanna McAdam and Paola Dey

Stroke. published online November 20, 2012;
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/early/2012/11/20/STROKEAHA.112.671743

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org/subscriptions/