Appendix A

We prove Theorem 2 by first establishing a lemma.

**Lemma 1.** Let

\[
\hat{\gamma}_{\text{mlt},n} = \left( \frac{n}{2} \right)^{-1} \sum_{(i,j) \in C_2^n} v_{ij\text{mlt}}, \quad \hat{\gamma}_{\text{t},n} = \left( \frac{n}{2} \right)^{-1} \sum_{(i,j) \in C_2^n} v_{ij\text{t}}, \quad \hat{\gamma}_{\text{t},n} = \left( \hat{\gamma}^\top_{12\text{t},n}, \hat{\gamma}^\top_{13\text{t},n}, \ldots, \hat{\gamma}^\top_{(M-1)\text{t},n} \right)^\top,
\]

\[
\hat{\gamma}_n = \left( \frac{n}{2} \right)^{-1} \sum_{(i,j) \in C_2^n} v_{ij} = \left( \hat{\gamma}^\top_{1\text{n}}, \hat{\gamma}^\top_{2\text{n}}, \ldots, \hat{\gamma}^\top_{T\text{n}} \right)^\top.
\]

Then, \( E(\hat{\gamma}_n) = 0 \) and

\[
\sqrt{n} \hat{\gamma}_n = \sqrt{n} \left( \frac{n}{2} \right)^{-1} \sum_{(i,j) \in C_2^n} v_{ij} \rightarrow_d N\left( 0, \Sigma_\gamma = 4\text{Var}(\hat{\nu}_i) \right).
\]  \( \quad (A.1) \)

We first prove Lemma 1.

**Proof.** It is readily checked by the iterated conditional expectation that

\[
E(\hat{\gamma}_{\text{mlt},n}) = E(v_{ij\text{mlt}}) = E\left[ E\left( \Delta^{-1}_{\text{mlt}} \Delta^{-1}_{\text{jmlt}} r_{\text{imlt}} r_{\text{ilt}} r_{\text{jmt}} r_{\text{jlt}} (u_{ij\text{mlt}} - \zeta_{\text{mlt}}) \mid y_i, y_j \right) \right] \quad (A.2)
\]

\[
= E\left[ \Delta^{-1}_{\text{mlt}} \Delta^{-1}_{\text{jmlt}} (u_{ij\text{mlt}} - \zeta_{\text{mlt}}) E(r_{\text{imlt}} r_{\text{ilt}} r_{\text{jmt}} r_{\text{jlt}} \mid y_i, y_j) \right]
\]

\[
= E\left[ \Delta^{-1}_{\text{mlt}} \Delta^{-1}_{\text{jmlt}} (u_{ij\text{mlt}} - \zeta_{\text{mlt}}) E(r_{\text{imlt}} r_{\text{ilt}} \mid y_i) E(r_{\text{jmt}} r_{\text{jlt}} \mid y_j) \right]
\]

\[
= E\left[ \Delta^{-1}_{\text{mlt}} \Delta^{-1}_{\text{jmlt}} \Delta_{\text{imlt}} \Delta_{\text{jmlt}} (u_{ij\text{mlt}} - \zeta_{\text{mlt}}) \right]
\]

\[
= 0.
\]
It follows that $E(\tilde{\gamma}_n) = 0$. Since $v_{jkmlt} = v_{jkmlt} (j \neq k)$, we have

$$\tilde{v}_{imlt} = E(v_{jkmlt} | y_i, r_i) = \begin{cases} 0 & \text{if } j \neq i, k \neq i \\ E(v_{ikmlt} | y_i, r_i) & \text{if } j = i \\ E(v_{ijmlt} | y_i, r_i) & \text{if } k = i \end{cases}.$$ 

Let $\Upsilon(i) = \{(j,k) \in C_n^2; j \neq i, k \neq i\}$. It then follows that

$$E(\tilde{\gamma}_{mlt,n} | y_i, r_i) = \left(\frac{n}{2}\right)^{-1} \left[ \sum_{(j,k) \notin \Upsilon(i)} E(v_{jkmlt} | y_i, r_i) + \sum_{(j,k) \in \Upsilon(i)} E(v_{jkmlt} | y_i, r_i) \right]$$

$$= \left(\frac{n}{2}\right)^{-1} \left[ \sum_{(j,k) \notin \Upsilon(i)} E(v_{jkmlt} | y_i, r_i) \right]$$

$$= \left(\frac{n}{2}\right)^{-1} \left[ \sum_{j=i,(j,k) \notin \Upsilon(i)} E(v_{jkmlt} | y_i, r_i) + \sum_{k=i,(j,k) \notin \Upsilon(i)} E(v_{jkmlt} | y_i, r_i) \right]$$

$$= \left(\frac{n}{2}\right)^{-1} \left[ \sum_{k=i+1}^{n} E(v_{ikmlt} | y_i, r_i) + \sum_{j=1}^{i-1} E(v_{ijmlt} | y_i, r_i) \right]$$

$$= \frac{2}{n} E(v_{ijklt} | y_i, r_i) = \frac{2}{n} \tilde{v}_{imlt}.$$ 

Thus, the projection of $\tilde{\gamma}_{mlt,n}$ is given by (e.g. Serfling, 1980, Chap.5):

$$\tilde{\tilde{\gamma}}_{mlt,n} = \sum_{i=1}^{n} E(\tilde{\gamma}_{mlt,n} | y_i, r_i) = \frac{1}{n} \sum_{i=1}^{n} 2\tilde{v}_{imlt}.$$ 

Since $\tilde{\gamma}_{mlt,n}$ is a sum of independently and identically distributed random variables, it follows from the central limit theorem (CLT) that:

$$\sqrt{n}\tilde{\gamma}_{mlt,n} = \frac{\sqrt{n}}{n} \sum_{i=1}^{n} 2\tilde{v}_{imlt} \rightarrow_d N(0, \Sigma_{mlt} = 4\text{Var}(\tilde{v}_{imlt})).$$

By the U-statistics theory (e.g. Serfling, 1980, Chap.5), $\tilde{\gamma}_{mlt,n}$ and $\hat{\gamma}_{mlt,n}$ have the same asymptotic distribution and thus:

$$\sqrt{n}\tilde{\gamma}_{mlt,n} \rightarrow_d N(0, \Sigma_{mlt} = 4\text{Var}(\tilde{v}_{imlt})).$$

The lemma follows by applying a similar argument to the vector $\tilde{\gamma}_n$. 

2
We now prove theorem 2.

Proof. By an argument similar to (A.2), we have:

\[ E \left( \frac{r_{imtlt}r_{jmtlt}}{\Delta_{imlt}\Delta_{jmlt}} u_{ijmlt} \right) = \zeta_{mlt}. \]

It then follows from the theory of U-statistics that

\[ \hat{\zeta}_{mlt} = \left( \frac{n}{2} \right)^{-1} \sum_{(i,j) \in C_2^n} \frac{r_{imtlt}r_{jmtlt}}{\Delta_{imlt}\Delta_{jmlt}} u_{ijmlt} \to_p \zeta_{mlt}. \]

Thus, by Slutsky’s theorem, \( \hat{\kappa}_{mlt} \) is a consistent estimate. Further, by applying the Lemma and Slutsky’s theorem, we obtain the asymptotic distribution of \( \hat{\kappa}_{mlt} \):

\[ \sqrt{n} \left( \hat{\zeta}_{mlt} - \zeta_{mlt} \right) = \sqrt{n} \left( \frac{n}{2} \right)^{-1} \sum_{(i,j) \in C_2^n} \hat{\gamma}_{mlt,n} \to_d N \left( 0, 4 \text{Var} \left( \tilde{v}_{imlt} \right) \right). \]

Similarly, by considering the vector \( \hat{\zeta} \), we obtain:

\[ \sqrt{n} \left( \hat{\zeta} - \zeta \right) \to_d N \left( 0, \Sigma_{\zeta} = 4 \text{Var} \left( \tilde{v}_i \right) \right). \]

Theorem 2 follows by applying the Delta method to \( \hat{\kappa} = f \left( \hat{\zeta} \right) \).

To show (13), first note that

\[ E \left( u_{ijmlt1} \mid y_i \right) = \frac{1}{2} E \left[ \sum_{g \in P^2} w_g \left( y_{imtg} - y_{jmtg} \right) (y_{iltg} - y_{jltg}) \mid y_i \right] \]

\[ = \frac{1}{2} \sum_{g \in P^2} w_g \left( y_{imtg} y_{iltg} - y_{jmtg} \phi_{iltg} - \phi_{mtg} y_{iltg} + \pi_{mtg} \right), \]

\[ E \left( u_{ijmlt2} \mid y_i \right) = 1 - \frac{1}{2} E \left[ \sum_{g \in P^2} w_g \left( y_{imtg} y_{jltg} + y_{iltg} y_{jmtg} \right) \mid y_i \right] \]

\[ = 1 - \frac{1}{2} \sum_{g \in P^2} w_g \left( y_{imtg} \phi_{ltg} + y_{iltg} \phi_{mtg} \right), \]
confirming the identities in (6). Further, we have:

\[ \tilde{v}_{imlt} = E (v_{ijmlt} | y_i, r_i) \]

\[ = E \left[ \Delta_{imlt}^{-1} \Delta_{jmlt}^{-1} r_{imtr_{ilt} r_{jmltr_{jlt}} (u_{ijmlt} - \zeta_{mlt})} | y_i, r_i \right] \]

\[ = \Delta_{imlt}^{-1} \Delta_{jmlt}^{-1} r_{imtr_{ilt}} E \left[ \Delta_{jmlt}^{-1} r_{jmltr_{jlt}} (u_{ijmlt} - \zeta_{mlt}) | y_i, y_j, r_i \right] \]

\[ = \Delta_{imlt}^{-1} \Delta_{jmlt}^{-1} r_{imtr_{ilt}} E \left[ \Delta_{jmlt}^{-1} (u_{ijmlt} - \zeta_{mlt}) E \left[ r_{jmltr_{jlt}} | y_i, y_j, r_i \right] | y_i, r_i \right] \]

\[ = \Delta_{imlt}^{-1} \Delta_{jmlt}^{-1} r_{imtr_{ilt}} E \left[ \Delta_{jmlt}^{-1} r_{jmltr_{jlt}} (u_{ijmlt} - \zeta_{mlt}) | y_i, r_i \right] \]

\[ = \Delta_{imlt}^{-1} \Delta_{jmlt}^{-1} r_{imtr_{ilt}} E \left[ (u_{ijmlt} - \zeta_{mlt}) \right]. \]

Thus, (13) follows by taking the covariances between \( \tilde{v}_{imlt} \) and \( \tilde{v}_{imlr_{ll'}} \) and substituting consistent estimates of \( E (u_{ijmlt} | y_i) \) using the identities in (6) with consistent estimates of \( \pi_{mltg} \), \( \phi_{mtgm} \) and \( \phi_{ltgl} \) given in (12).

**Appendix B**

We show that it is not possible to model the missingness of \( y_{i1t} \) as dependent on \( y_{i2t} \) and vice versa in addition to \( H_{it} \) under MAR. For convenience, we consider the case with \( t = 1 \) and denote \( y_{i1t} (y_{i2t}) \) simply as \( y_{i1} (y_{i2}) \). For notational brevity, we also suppress the dependence on \( H_{it} \).

Suppose that on the contrary such a model existed. Then, we would have:

\[ \Pr [r_{i1} = 1 \mid r_{i2} = 1, y_{i1}, y_{i2}] = \Pr [r_{i1} = 1 \mid r_{i2} = 1, y_{i2}], \]  

\[ \Pr [r_{i2} = 1 \mid r_{i1} = 1, y_{i1}, y_{i2}] = \Pr [r_{i2} = 1 \mid r_{i1} = 1, y_{i1}]. \]  

Under MAR, the probabilities of missing response only depend on observed data. It follows that \( \Pr [r_{i1} = 0, r_{i2} = 0 \mid y_{i1}, y_{i2}] \) is a constant, say \( c \) (0 < c < 1), \( \Pr [r_{i1} = 1, r_{i2} = 0 \mid y_{i1}, y_{i2}] \)
is a function of $y_{i1}$ and $\Pr [r_{i1} = 0, r_{i2} = 1 \mid y_{i1}, y_{i2}]$ a function of $y_{i2}$ only. Denote the latter two quantities as $f(y_{i1})$ and $g(y_{i2})$, respectively. Then, $\Pr [r_{i1} = 1, r_{i2} = 1 \mid y_{i1}, y_{i2}] = 1 - f(y_{i1}) - g(y_{i2}) - c$. It follows that

\[
\Pr [r_{i1} = 1 \mid y_{i1}, y_{i2}] = f(y_{i1}) + 1 - f(y_{i1}) - g(y_{i2}) - c = 1 - g(y_{i2}) - c,
\]

\[
\Pr [r_{i2} = 1 \mid r_{i1} = 1, y_{i1}, y_{i2}] = \frac{1 - f(y_{i1}) - g(y_{i2}) - c}{1 - g(y_{i2}) - c}.
\]

It follows from (B.3) that $\frac{1-f(y_{i1})-g(y_{i2})-c}{1-g(y_{i2})-c}$ is a function of $y_{i1}$ only. Thus, $g(y_{i2})$ must be a constant. Likewise, $f(y_{i1})$ must be a constant. These contradict the MAR assumption.

APPENDIX C

C0.1 Simulation method

We conducted a limited simulation study to examine the empirical size of type I error rate for testing the null of equal kappas between two raters under a longitudinal study design with three assessment times for four sample sizes: 30, 50, 75 and 100. For each of the sample sizes under study, we generated pairs of observers’ ratings from a three-level categorical response at three assessment times, $y_{it} = (y_{i1t}^\top, y_{i2t}^\top)^\top$ ($1 \leq t \leq 3$), under the null, $H_0 : k_1 = k_2 = k_3$, with $k_t$ denoting the between-rater kappa at time $t$. We tested the null using the $\chi^2$ statistic $Q^2_n$ and estimated the type I error rate based on the empirical distribution of the test statistic obtained from 1,000 Monte Carlo (MC) replications.

We created the pairs of ratings $y_{it}$ over three assessment times from a three-level categorical outcome in two steps.

Step 1. We generated pairs of continuous outcomes over three assessment times, $z_{it} = (z_{i1t}, z_{i2t})^\top$, by simulating a six-dimensional vector, $z_i = (z_{i11}, z_{i21}, z_{i12}, z_{i22}, z_{i13}, z_{i23})^\top$, from a six-variate normal distribution with mean vector $\mathbf{0}$ and variance $\Sigma = C(\rho)$, with $C(\rho)$ denoting a compound symmetry correlation matrix with the correlation set at $\rho = 0.5$. 

5
Step 2. We transformed each of the continuous observer score $z_{imt}$ into a three-level categorical response by:

$$y_{imt} = \left( I\{z_{imt} \leq \Phi^{-1}(1/3)\}, I\{\Phi^{-1}(1/3) < z_{imt} \leq \Phi^{-1}(2/3)\}, I\{z_{imt} > \Phi^{-1}(2/3)\} \right)^\top,$$

where $\Phi^{-1}$ denotes the inverse of the CDF of the standard normal.

For a given sample size, let $Q_j$ be the test statistic from the $j$th MC replication and $G_{0.95}$ the 95th percentile of the $\chi^2$ distribution with 2 degrees of freedom. The empirical size is calculated as $\hat{\alpha} = \frac{1}{1000} \sum_{j=1}^{1000} I\{Q_j \geq G_{0.95}\}$.

For the missing data case, we assumed no missing data at baseline $t = 1$ and simulated the missing response according to a MCAR and MAR model.

For MAR, we considered the MMDP model with lag time $L = 1$ and simulated the missing data indicators for the two raters at each time $t$, $(r_{i1t}, r_{i2t})$, according to a multinomial model with the probability vector $p_{it} = (p_{i21}, p_{i31}, p_{i41})^\top$ ($t = 2, 3$). We specified the one-step transition probabilities $p_{ilt}$ at times $t$ according to the generalized logit model in (16). For convenience, we modeled the dependence on the response under the Markov assumption by converting the three-level categorical outcome $y_{imt}$ to a continuous scale $y_{imt}$ with three values, 1, 2 and 3, i.e., $y_{imt} = (1, 2, 3)^\top y_{imt}$. To have about 10% and 15% missing responses at times 2 and 3, we solved the following equations for $\beta_{2t}$ and $\beta_{3t}$:

$$0.9n = \sum_{i=1}^{n} p_{i21}, \quad 0.85n = \sum_{i=1}^{n} p_{i21} p_{i31}, \quad (C.4)$$

where

$$p_{i21} = \frac{1}{1 + \sum_{l=2}^{4} \exp (\beta_{2l} (y_{i11} + y_{i21}))}, \quad p_{i31} = \frac{1}{1 + \sum_{l=2}^{4} \exp (\beta_{3l} (y_{i12} + y_{i22}))}. \quad (C.5)$$

To ensure that the missing data indicators $(r_{i1t}, r_{i2t})$ follow the MMDP model with $L = 1$, we further imposed the following restrictions:

$$r_{i13} = r_{i12} \times r_{i22} \times r_{i13}, \quad r_{i23} = r_{i12} \times r_{i22} \times r_{i23}.$$
For MCAR, the same approach above was used except that $p_{i21}$ and $p_{i31}$ were modeled independently of $y_{imt}$ with $p_{i21} = 0.90$ and $p_{i31} = \frac{0.85}{0.90}$ to produce about 10% and 15% missing responses at $t = 2, 3$, respectively.

C0.2 Simulation results

As shown in the above table, the empirical size varied from 0.078 to 0.06 for the complete, and from 0.086 (0.087) to 0.064 (0.07) for the missing data case under MCAR (MAR) as the sample size increased from 30 to 100. The empirical sizes are slightly larger than the nominal size 0.05, with the bias increasing in order when moving from the complete data case to MCAR and to MAR. As expected, the empirical size becomes closer to 0.05 in all cases as the sample size increases.

REFERENCES

Table 1
The empirical size of type I error rate for testing the null of equal kappas between two raters under a longitudinal study design with three assessment times for four sample sizes: 30, 50, 75 and 100.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Complete</th>
<th>MCAR</th>
<th>MAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.078</td>
<td>0.079</td>
<td>0.081</td>
</tr>
<tr>
<td>50</td>
<td>0.077</td>
<td>0.086</td>
<td>0.087</td>
</tr>
<tr>
<td>75</td>
<td>0.060</td>
<td>0.065</td>
<td>0.071</td>
</tr>
<tr>
<td>100</td>
<td>0.064</td>
<td>0.064</td>
<td>0.070</td>
</tr>
</tbody>
</table>