**SDC 2: Level of evidence**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Timed Up and Go test protocol** |  | **Type of reliability** |  | **Number of studies** | **+** | **Methodological quality: COSMIN** | **+** | **Consistency of findings** | **=** | **Level of evidence** |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Williams et al. (2005)10 |  | Intra-rater Test-retest |  | 1 | + | Excellent | + | Consistent | = | Strong |
|  |
|  |  |  |  | ≥ 2 | + | Good | + | Consistent |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Nicolini-Panisson & Donadio (2014)6 |  | Intra-rater Test-retest |  | 1 | + | Good | + | Consistent | = | Moderate |
|  |  |  | ≥ 2 | + | Fair | + | Consistent |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | 1 | + | Fair | + | Not applicable | = | Limited |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Butz et al. (2015)23 |  | Intra-rater |  | ≥1 | + | Poor | + | Consistent | = | Unknown |
| Habib et al. (1999)8 |
|  |  |  |
| Butz et al. (2015)23 |  | Inter-rater |
| Habib et al. (1999)8 |
| Itzowitz et al. (2016)24 |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | ≥1 | + | Poor - excellent | + | Inconsistent | = | Conflicting |
| Legend: Relative reliability refers to the consistency of TUG time within a session (intra-rater reliability), between raters (inter-rater reliability) and between test sessions (test-retest reliability). |
|  |  |  |

 **2A**: This flowchart depicts the level of evidence assignment for relative reliability. The criteria of the Cochrane Collaboration Back Review Group (van Tulder et al. 200321) were used to determine the level of evidence, but implemented as by Saether et al. 201322.

**Relative reliability**

**2B:** This flowchart depicts the level of evidence assignment for absolute reliability. The criteria of the Cochrane Collaboration Back Review Group (van Tulder et al. 200321) were used to determine the level of evidence, but implemented as by Saether et al. 201322.

**Absolute reliability**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Timed Up and Go test protocol** |  | **Type of reliability** |  | **Number of studies** | **+** | **Methodological quality: COSMIN** | **+** | **Consistency of findings** | **=** | **Level of evidence** |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Williams et al. (2005)10 |  | Intra-rater |  | 1 | + | Excellent | + | Consistent | = | Strong |
| Nicolini-Panisson & Donadio (2014)6 | Test-retest |  |
|  |  |  |  | ≥ 2 | + | Good | + | Consistent |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | 1 | + | Good | + | Consistent | = | Moderate |
|  |  |  | ≥ 2 | + | Fair | + | Consistent |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | 1 | + | Fair | + | Not applicable | = | Limited |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | ≥1 | + | Poor | + | Consistent | = | Unknown |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | ≥1 | + | Poor - excellent | + | Inconsistent | = | Conflicting |
| Legend: Absolute reliability refers to the measurement error within a session (intra-rater reliability) and between test sessions (test-retest reliability). |
|  |  |  |