Intraclass Correlation Coefficients in Multilevel Structural Validation of Intensively Collected Questionnaires: A Targeted Literature Review

Christina Daskalopoulou[1], Aleksandra Sjöström-Bujacz[2], Jakob Bjørner[3], Dara O'Neill[4]

[1]IQVIA, Athens, Greece; [2]IQVIA, Stockholm, Sweden; [3]IQVIA, Copenhagen, Denmark; [4]IQVIA, Barcelona, Spain

ISOQOL 32nd Annual Conference, 22nd-25th October, Milwaukee, Wisconsin, USA

Background

- The past decade has seen an **Increase in Intensive Longitudinal Data (ILD) collection** through Clinical Outcome Assessments (COAs), such as Daily Diaries (DD) and digital health measures.
- ILD enables the examination of both **within-person (intraindividual) and between-person (interindividual) variability** in health and psychological constructs.
- In ILD, COA scores may vary substantially between assessments (e.g., day to day, moment to moment), yet such within-person variation is overlooked by **traditional single-level structural validity models**.
- **Multilevel confirmatory factor analysis (MCFA)** allows researchers to model latent structures separately at the within- and between-person level, improving insights into the performance of a scoring algorithm for multi-item questionnaires.

Objectives

- To conduct a **Targeted Literature Review (TLR)** of ILD studies that applied multilevel modeling framework to COAs.
- To examine the use of **Intraclass Correlation Coefficients (ICCs)** as indicators of between-person variability, and the use of (**1-ICCs**) as indicators of within-person variability.
- To assess **how ICC values inform the need for multilevel structural modeling** strategies, such as MCFA.
- To identify gaps in reporting practices and propose recommendations for **improving transparency** and **standardization** in future research.

Methods

- PsycNet, PsycInfo, PubMed, and Web of Science were searched on November 11, 2024. Search terms included 'intraclass correlation coefficient' together with 'ecological momentary assessment (EMA)', 'experience sampling', 'digital', 'diary', 'ambulatory assessments' and 'multilevel', 'hierarchical' structural validity. No date restrictions were applied. Additional papers were identified through snowballing.
- Two independent reviewers sequentially performed title/abstract screening followed by full text screening for each reference. Disagreements were resolved through discussion and additional review by a third independent reviewer.
- **Figure 1** presents the PRISMA flow chart of the screening and selection process.

Figure 1. PRISMA flow chart
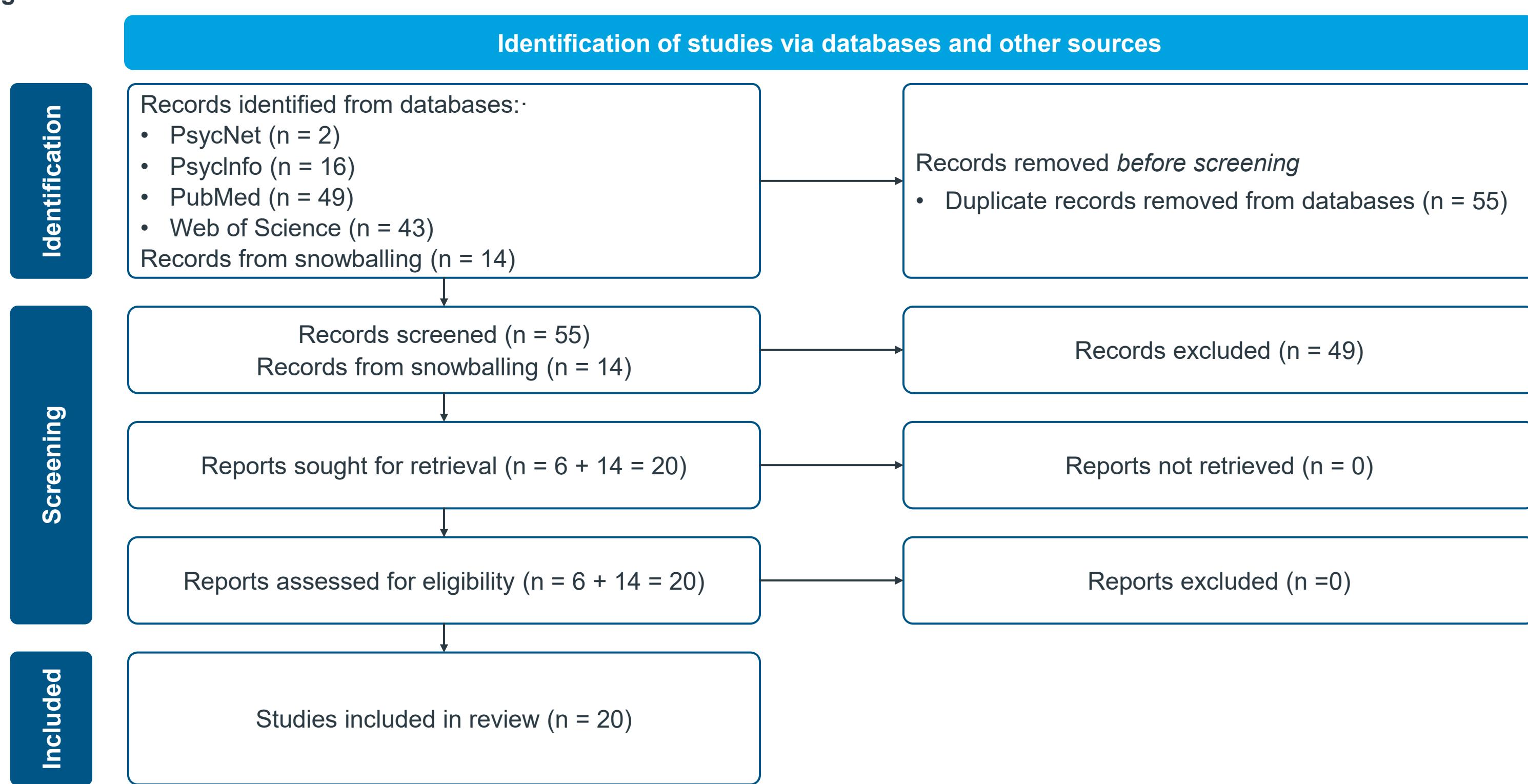
Source: Page MJ, et al. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

Results

- 20 studies were included, with sample sizes ranging from 49 to 2,104 and total observations, defined as participants × number of assessments, ranging from 377 to 29,950 (median: 1,026). 11 studies reported both sample sizes and total observations, while 9 studies reported only the sample size.
- Between-person variability (ICCs), representing the proportion of total variance attributable to between-person differences, was notably heterogenous across studies. Studies that utilized observational DD designs and/or concerned chronic physical health conditions had higher average maximum ICC values, indicating greater between-person variability (**Table 1a** and **1b**). In contrast, studies that used interventional EMA designs and/or investigated mental health and emotional functioning had lower ICC values, suggesting more variability in participants responses across time than among them.
- There has been an increase in the number of observational DD and observational EMA studies in the last 5 years, and in those concerning mental health and emotional functioning (**Figure 2a** and **2b**).
- In 16 of 20 studies, the same number of factors was identified at both the within- and between-person levels. In 3 studies, the number of factors differed, and 1 study assessed only the within-person level factor structure.

Table 1a. Summary of average ICC values per study design

Prospective study design	Number of studies	Average of the min ICC values	Average of the max ICC values
Observational (DD)	11	0.331	0.682
Observational (EMA)	4	0.338	0.571
Interventional (DD)	1	0.170	0.620
Interventional (EMA)	1	0.090	0.317
Observational (DD and EMA)	2	0.214	0.535
Overall	20	0.301	0.618

DD: Daily Diaries; EMA: Ecological Monetary Assessment; ICC: Intraclass Correlation Coefficient at the between level referring to the proportion of total variance across observed items attributable to differences between persons.

Table 1b. Summary of average ICC values per therapeutic area

Therapeutic area category	Number of studies	Average of the min ICC values	Average of the max ICC values
Mental health & emotional functioning	9	0.253	0.541
Occupational stress & fatigue	4	0.315	0.615
Chronic physical health conditions	4	0.438	0.785
Substance use & addiction	3	0.270	0.630
Overall	20	0.301	0.618

Figure 2a. Number of studies per study design and publication year

Figure 2b. Number of studies per therapeutic area and publication year
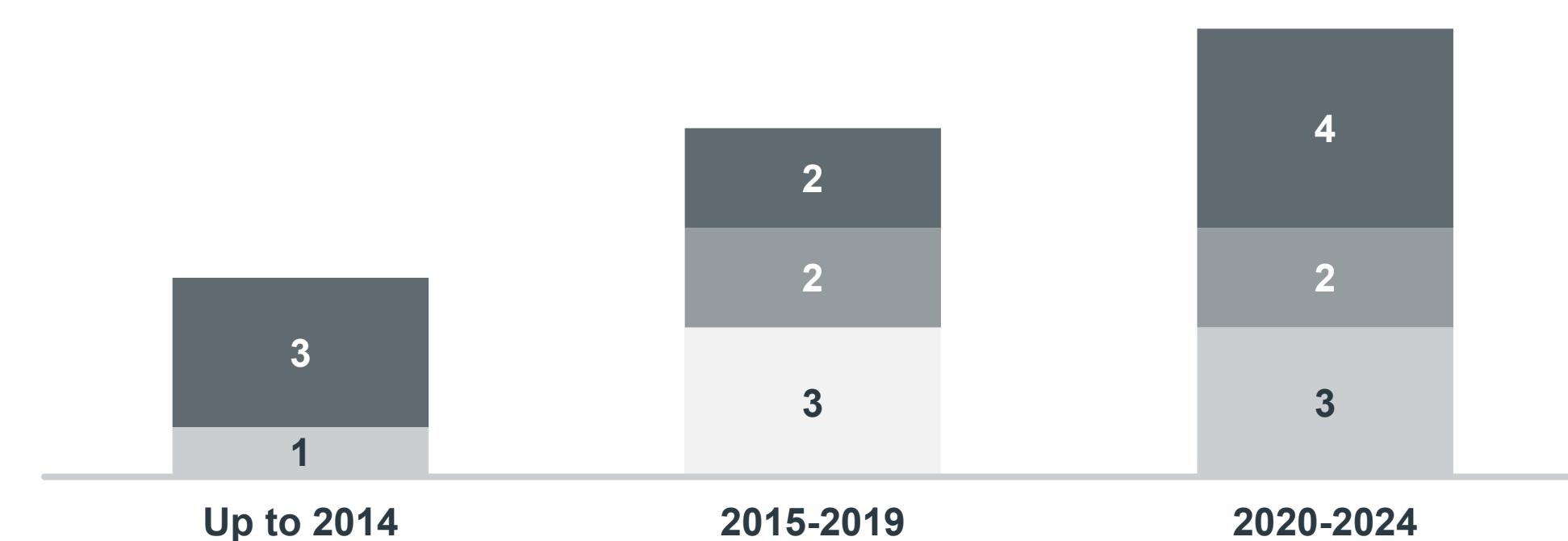
Conclusions

- **MCFA**, revealing whether constructs are assessed similarly within and between persons, **is increasingly applied** in ILD studies using COAs, especially in those including COAs assessing mental health and emotional functioning and chronic physical health conditions.
- Our TLR showed that **ICC values vary across study design and therapeutic area, highlighting differences in within- and between-person variability.** Higher ICCs in observational DD design and chronic physical health measures suggest more between-person variability, while lower ICCs in interventional EMA designs and mental health measures highlight the importance of capturing within-person variability.
- Finally, standardized reporting of modeling strategies in publications, abstracts, and keywords is needed to improve **transparency, comparability,** and **discoverability** in future research.

References

[1]Alameer et al., 2023, doi.org/10.1037/ocp0000354; [2]Dumi et al., 2024, doi.org/10.1080/10543406.2024.2310312; [3]Heshmati et al., 2023, doi.org/10.1007/s10902-023-00684-w; [4]Jacobson et al., 2023, doi.org/10.1037/emo0000102; [5]Kehayes & Mackinnon, 2019, doi.org/10.1525/collabra.257; [6]Klootwijk et al., 2021; doi.org/10.1111/jora.12660; [7]Lambusch et al., 2020; doi.org/10.1007/978-3-030-49044-7_16; [8]Lee et al., 2015, doi.org/10.15288/jsad.2015.76.326; [9]Lee et al., 2017, doi.org/10.1037/pas0000320; [10]Mason et al., 2024, doi.org/10.1177/01632787241249500; [11]McMahon & Naragon-Gainey, 2019, doi.org/10.1177/2167702618807408; [12]Merz & Roesch, 2011, doi.org/10.1016/j.jrp.2010.11.003; [13]Reise et al., 2005, doi.org/10.1207/s15327752jpa8402_02; [14]Riedl & Thomas, 2019, doi.org/10.1080/1359432X.2019.1588251; [15]Roesch et al., 2010, doi.org/10.1080/00273171.2010.519276; [16]Sirbu et al., 2022, doi.org/10.3389/fneur.2022.941788; [17]Saholm et al., 2023, doi.org/10.1371/journal.pone.0283148; [18]Weigelt et al., 2019, doi.org/10.1037/ocp0000117; [19]Wenze et al., 2018, doi.org/10.1016/j.brat.2018.05.011; [20]Wolff et al., 2012, doi.org/10.1037/a0029125