WHITE PAPER AUGUST 2026

Lost in Migration

When paper instruments break on screens — and how to move them faithfully

IN THIS PAPER

ECOALA PTY LTD · AUSTRALIA
ECOALA.COM.AU

© 2026 ECOALA PTY LTD

ABSTRACT

Most clinical outcome assessment (COA) instruments in use today were designed for paper, and most trials now administer them on screens. That migration is usually treated as an administrative step; it is in fact a measurement intervention. A recall period rendered differently, a response scale that behaves differently under a fingertip than a pen, a page break that changes item context — each can alter what the instrument measures, and none is visible in the resulting dataset. This paper reviews the peer-reviewed evidence of the past decade on paper-to-electronic migration: where equivalence has been convincingly demonstrated, where it has not, what the updated ISPOR good-practice framework actually requires of sponsors, and how the discipline of faithful migration — design rules, usability evidence, dataset standards and now AI-assisted processes — protects data quality. It is written for clinical trial and pharmaceutical professionals who work with eCOA outputs without being eCOA specialists.

01 Migration is a measurement decision, not an IT task

When a validated paper questionnaire is rebuilt for a smartphone or tablet, everything about its measurement pedigree — its validation studies, its interpretation thresholds, its regulatory history — was earned in a different medium. The industry's governing assumption is that a faithful migration preserves those properties. The assumption is usually right, but 'faithful' is doing enormous work in that sentence, and the failure modes are quiet: a dataset produced by a subtly altered instrument looks exactly like a dataset produced by the original.

The stakes have grown with adoption. Electronic capture is now the default in industry trials, encouraged by regulators for its completeness, timestamps and audit trails, and extended by bring-your-own-device (BYOD) designs onto hardware the sponsor has never seen[1,2]. Every one of those advantages is real; none of them substitutes for measurement comparability.

© 2026 ECOALA PTY LTD

02 What can actually change on a screen

The mechanics of screen administration differ from paper in specific, studyable ways. Items that share a page on paper may be presented one per screen; instructions and recall periods may scroll out of view; response scales — visual analogue scales in particular — are re-rendered at different physical sizes; navigation rules (forced responses, no backtracking) alter how respondents engage; and on BYOD, screen size, operating system and font rendering vary across every participant in the study[2,3]. Encouragingly, the empirical record shows that well-executed migrations survive these changes. A randomised three-way crossover trial comparing paper, a provisioned device and participants' own devices found high agreement across all widely used response scale types — visual analogue, numeric rating, verbal response and Likert — with intraclass correlation coefficients (ICCs) between 0.79 and 0.98, alongside strong patient preference for electronic completion[3]. A dedicated study of scrolling — long considered a threat because respondents may answer before seeing all information — found ICCs of 0.71 to 0.96 between scrolling and non-scrolling presentations, with the equivalence threshold met for all but three of the scores examined[4]. A review of the accumulated equivalence literature concluded that the evidence overwhelmingly supports measurement equivalence of instruments migrated to screen-based formats when best-practice design rules are followed, including in BYOD settings where a minimum device specification can be assured[2].

The important caveat sits inside that conclusion: equivalence is conditional on best practice. The same body of work documents that the marginal cases — the item scores that fell short of the equivalence threshold in the scrolling study, for example[4] — cluster exactly where design discipline lapses. Equivalence is the reward for faithful migration, not a property of screens.

03 What the current good-practice framework requires

The governing methodological reference is the 2023 ISPOR task force report on measurement comparability among modes of data collection, which updated the 2009 and 2014 guidance in light of the accumulated evidence[1]. Its central logic is evidence-based proportionality: sponsors should first determine whether existing comparability evidence covers the questionnaire and the technology in question. Where sufficient evidence exists and best practices for faithful migration are followed, the task force concludes that further comparability testing is unnecessary — including for mixed-modes and BYOD designs[1].

© 2026 ECOALA PTY LTD

Where the migration is not faithful — where wording, response options, recall periods or instructions change — the migration becomes a modification, and the evidentiary burden escalates accordingly.

Two practical implications follow for non-specialists. First, the equivalence question is answered at the level of the migration approach and component types, not necessarily per study: a standard instrument, migrated under standard rules, onto standard response scale types, generally inherits the existing evidence base[1,2]. Second, the sponsor's real obligations shift upstream, into documented design fidelity and usability evidence in a representative population — which is precisely where corners get cut under timeline pressure.

04 The populations that break the assumptions

Averages conceal the users for whom screens are genuinely harder. Qualitative work with cancer patients found that peripheral neuropathy of the hands, fatigue, concentration and memory problems, and position within a treatment cycle all affected participants' ability to interact with ePRO solutions; the recommended mitigations — larger and well-spaced buttons, the ability to pause and resume, and re-presenting the recall period with every question — are simple, but only if elicited before the build rather than discovered after database lock[5]. Migration programmes that skip representative usability work import an invisible mode effect concentrated in the sickest participants — often the very participants whose data carry the endpoint.

Clinician-administered instruments raise their own mode questions as trials decentralise. A multicentre randomised comparison of remote versus in-person administration of the Montgomery– Åsberg Depression Rating Scale in major depressive disorder found strong agreement (overall ICC 0.886), supporting remote administration as a feasible alternative — while also observing that consistency decreased as depression severity increased, a reminder that mode effects can be severity-dependent even when the average looks reassuring[6].

05 After the screen: data structure as part of fidelity

Faithful migration does not end at the user interface. A multistakeholder best-practice initiative on ePRO dataset structure documented that electronic COA data are not required to follow any standard data model, that models vary by provider and sponsor, and that this inconsistency creates risks for programming, analysis and regulatory

© 2026 ECOALA PTY LTD

submission; its recommendations — adopting CDISC standards within the eCOA platform, early stakeholder involvement, planned handling of missing data, and validated, quality-controlled datasets — treat the dataset itself as part of the instrument[7]. A migration that preserves every pixel of the questionnaire but scrambles the provenance, versioning or structure of its data has still broken the instrument, just further downstream.

The newest frontier is the use of artificial intelligence within migration and translation workflows. Recent good-practice work by an ISOQOL special interest group — combining a literature review, a fifty-stakeholder survey and expert interviews — offers the first structured recommendations for where AI can and cannot appropriately assist COA linguistic validation and eCOA migration, balancing efficiency gains against intellectual-property, data-privacy and patient-centricity concerns[8]. The direction is clear: AI will compress migration timelines, and the governance question is how to let it do so without eroding exactly the fidelity this paper has been describing.

06 A migration checklist for non-specialists

For sponsors and study teams who consume rather than produce eCOA science, the accumulated evidence supports a short set of questions to ask of any migration. Is the migration faithful — identical wording, response options, recall periods and instructions — and is that fidelity documented item by item[1]? Do the response scale types used carry existing equivalence evidence, or does the instrument include unusual components that genuinely need new testing[2,3]? Has scrolling been either designed out or implemented with safeguards such as disabling progression until all content is viewed[4]? Has usability been tested in a representative sample of the actual trial population, including its most impaired members[5]? For remote or decentralised administration of clinician-rated measures, is there mode-comparability evidence for the instrument, and has severity dependence been considered[6]? And is the resulting dataset built to a defined standard with validation and quality control specified before first patient in[7]?

None of these questions requires specialist psychometrics to ask. All of them are cheaper to answer at design time than to litigate at submission.

07 Conclusion

The paper-to-digital transition has been one of the genuine data-quality successes of modern clinical research: the equivalence evidence is broad, the good-practice framework is mature, and the operational benefits are real[1,2,3]. But that success was built by treating

© 2026 ECOALA PTY LTD

migration as measurement science — design rules, equivalence studies, usability work, dataset standards — and it is only preserved by continuing to do so. Every instrument that arrives on a screen carries a validation history earned on paper. Whether it still deserves that history is decided, quietly, by the quality of its migration.

> **NOTE ON LITERATURE SOURCING**
>
> Literature for this white paper was identified through structured searches of PubMed (National Library of Medicine), restricted to peer-reviewed publications from 2016 onwards. Digital object identifiers (DOIs) are provided for all cited works.

References

01 O'Donohoe P, Reasner DS, Kovacs SM, Byrom B, Eremenco S, Barsdorf AI, et al. Updated recommendations on evidence needed to support measurement comparability among modes of data collection for patient-reported outcome measures: a good practices report of an ISPOR task force. Value in Health 2023;26(5):623–633. https://doi.org/10.1016/j.jval.2023.01.001

02 Byrom B, Gwaltney C, Slagle A, Gnanasakthy A, Muehlhausen W. Measurement equivalence of patient-reported outcome measures migrated to electronic formats: a review of evidence and recommendations for clinical trials and bring your own device. Therapeutic Innovation & Regulatory Science 2019;53(4):426–430. https://doi.org/10.1177/2168479018793369

03 Byrom B, Doll H, Muehlhausen W, Flood E, Cassedy C, McDowell B, et al. Measurement equivalence of patient-reported outcome measure response scale types collected using bring your own device compared to paper and a provisioned device: results of a randomized equivalence trial. Value in Health 2018;21(5):581–589. https://doi.org/10.1016/j.jval.2017.10.008

04 Shahraz S, Pham TP, Gibson M, De La Cruz M, Baara M, Karnik S, et al. Does scrolling affect measurement equivalence of electronic patient-reported outcome measures (ePROM)? Results of a quantitative equivalence study. Journal of Patient-Reported Outcomes 2021;5(1):23. https://doi.org/10.1186/s41687-021-00296-z

05 Mowlem FD, Sanderson B, Platko JV, Byrom B. Optimizing electronic capture of patient-reported outcome measures in oncology clinical trials: lessons learned from a qualitative study. Journal of Comparative Effectiveness Research 2020;9(17):1195–1204. https://doi.org/10.2217/cer-2020-0143

06 Sumiyoshi T, Morio Y, Kawashima T, Tachimori H, Hongo S, Kishimoto T, et al. Feasibility of remote interviews in assessing disease severity in patients with major depressive disorder: a pilot study. Neuropsychopharmacology Reports 2024;44(1):149–157. https://doi.org/10.1002/npr2.12411

07 Hudgens S, Kern S, Barsdorf AI, Cassells S, Rowe A, King-Kallimanis BL, et al. Best practice recommendations for electronic patient-reported outcome dataset structure and standardization to support drug development. Value in Health 2023;26(8):1242–1248. https://doi.org/10.1016/j.jval.2023.02.011

© 2026 ECOALA PTY LTD

08 McKown S, Arnold B, Boucher F, Correia H, Eremenco S, Geršić T, et al. Use of AI within COA linguistic validation and eCOA migration processes: analysis and good practice recommendations. Journal of Patient-Reported Outcomes 2026;10(1). https://doi.org/10.1186/s41687-026-01012-5

© 2026 ECOALA PTY LTD