Digital Parenting Research Toolkit: development of linked outcome-planning resources for studies of digital parent support

Development report and internal verification of version 0.3.0

Alexander Kuznetsov Singapore. Institutional affiliation and correspondence details require author confirmation. Manuscript draft: 23 October 2026. Author review pending; not peer reviewed or deposited as a preprint.

Abstract

Background: Research on digital programmes that support parents often combines caregiver and child reports, practitioner assessments and records of programme use. Reproducible measurement requires more than naming a questionnaire: the construct, exact instrument, information source, subjective rater, score calculation, timing and analysis must be connected. Missing and late assessments create additional record-management decisions.

Objective: To describe the development, scope, components and internal verification of the Digital Parenting Research Toolkit (DiPaR Toolkit), a free, openly licensed planning resource for this setting.

Methods: A targeted, non-systematic review of primary reporting guidance and instrument-provider information informed original prompts and templates. The principal use case was an individually randomised two-arm study with one index child per family. Version 0.3.0 links a browser project planner, guidance map, six candidate measurement cards, editable study records, an offline structural checker and a fictional teaching project. The planner distinguishes information provision, subjective judgment and deterministic calculation. Internal verification used reproducible software tests and deliberately missing, extra, invalid and late records. No researchers, caregivers or children were recruited for evaluation.

Results: The planner contains 37 fields, nine shared across outcomes, plus decision records and additional-occasion entries. A project exports linked definitions, a register and a schedule, with a blank independent-roster template. The fictional fixture contains two outcomes and 16 expected assessments from four children: 13 observed scores, two missing assessments and one unscorable partial response. Reconciliation produces no structural errors and one expected timing warning; a valid zero remains observed. Twelve JavaScript tests and 14 Python tests passed for the specified scenarios. The source, fixtures and commands accompany the release. These findings demonstrate selected software behaviours and content organisation, not usability, methodological completeness, clinical benefit or reduced risk of bias.

Conclusions: DiPaR Toolkit offers an inspectable way to connect measurement decisions and records for a narrow research use case. Independent methodological review, task-based user evaluation and study-specific measurement work are needed before claims of improved research quality can be made.

Keywords: digital parenting; parent support; outcome measurement; trial protocols; research software; missing data; reproducibility.

Introduction

Digital parent-support programmes differ in content, human support, access requirements and delivery. A programme may intend to change child symptoms, child functioning, parenting practices or parental wellbeing. These are distinct constructs and need separate justification. Software usage can describe

exposure but does not itself show benefit. Outcome collection may involve a parent describing a child's behaviour, a child describing an experience, or a clinician rating information supplied by a parent. These arrangements should remain visible in the study record.

SPIRIT 2025 gives recommendations for reporting randomised trial protocols [1]. SPIRIT-Outcomes adds detail about outcome definitions, assessment and analysis [2]. SPIRIT-C 2026 addresses paediatric considerations, including developmental differences, burden and age-appropriate measurement [3]. TIDieR supports intervention description [4]. COSMIN provides resources for selecting and evaluating measurement instruments [5]. SPIRIT-PRO is relevant where patient-reported outcomes or justified proxy reporting are involved; it should not be applied indiscriminately to every caregiver questionnaire [9,10]. CONSORT-EHEALTH adds digital-intervention reporting considerations to be considered alongside current CONSORT guidance [6,7,11].

Existing resources already support these tasks. The SPIRIT–CONSORT organisation describes SEPTRE as a protocol creation and management tool and COBWEB as a writing aid for randomised trial reports [12]. The PROTEUS Consortium provides a detailed SPIRIT-PRO protocol template [13]. DiPaR does not claim that protocol assistance is new or that these tools lack particular features. Its proposed contribution is a focused collection of linked records, instructional examples and structural checks for digital parent-support research. Whether that combination helps users is an empirical question.

Risk-of-bias assessment is a different task. RoB 2 examines a specified result, including the outcome and comparison, using its signalling questions and decision framework [8]. Completing DiPaR prompts cannot establish low risk of bias. Retaining a dated protocol, allocation records, expected assessment roster, measurement procedures and amendments may make later appraisal more transparent, but it does not prove that the planned procedures were followed.

This paper reports the rationale and implementation of a development release, the source-selection process, the executed internal checks and the remaining evidence gaps. It is not a trial report, validation study of a questionnaire, systematic review, consensus guideline or evaluation of toolkit effectiveness.

Methods

Scope and development process

The intended audience is individual researchers and small teams planning studies of digital programmes that help parents. The initial record structure assumes an individually randomised two-arm parallel study, one index child per family and outcome assessments at defined occasions. A feasibility design is supported as a planning option, while the user must specify feasibility outcomes and progression decisions separately. Cluster, multi-arm, adaptive and non-inferiority designs require additional work; selecting another design in the form does not add the necessary statistical or operational provisions.

The author established the intended scope and directed preparation of the resource. Generative AI assisted with content drafting, coding, source checks and internal verification. No formal stakeholder consultation, Delphi process, independent expert panel or systematic review of competing tools was conducted. Caregivers and children did not participate in development. The public-facing content and source files document these limitations rather than treating internal preparation as independent review.

Version 0.3.0 followed an internal audit of version 0.2.1. The audit identified a single-outcome workflow, manually linked files, an ambiguous distinction between the rater and score calculation, incomplete instructional decisions and a data checker that could not detect an entirely absent expected row. The revision introduced a project-level structure, explicit decision states and reconciliation against

independently supplied reference files. The earlier version is retained as an archive; new functionality is not retrospectively attributed to it.

Source identification and interpretation

The targeted source check, completed on 23 October 2026, used official guideline publications and project pages, instrument-provider or author pages, and selected original measurement studies. Sources were chosen for direct relevance to protocol specification, paediatric participation, reported outcomes, digital delivery or the candidate constructs. No comprehensive database search, screening flow, preregistered review protocol or quality appraisal of all measurement studies was undertaken. Consequently, the catalogue cannot be interpreted as an exhaustive evidence synthesis or ranking.

A developer location map relates the 37 planner fields and companion templates to selected guidance topics. It includes the 17 additional SPIRIT-C items and nine SPIRIT-Outcomes items, preserving original identifiers where the extension used SPIRIT 2013 numbering [2,3]. SPIRIT-PRO topics are mapped by applicability, including reporting source, assistance, missing assessments and clinical alerts [9,10,13]. Digital prompts cover programme version, human support, exposure definitions and changes [4,11]. The map is an interpretation: a prompt may cover only part of an item, and location in a template does not demonstrate compliance.

Six candidate cards were selected purposively to span child difficulties, anxiety/depression, disruptive behaviour, parental stress, parenting practices and anxiety-related functioning. The cards concern SDQ, RCADS, ECBI, PSI-4-SF, APQ and CAIS. Each identifies the form or edition under discussion, provider information, an original evidence anchor and unresolved suitability or rights questions [14–25]. Availability of an instrument or article is not treated as permission to reproduce questionnaire content. No items, scoring keys or normative tables are bundled.

Evidence is interpreted at the level of the relevant population, age range, language, form and measurement property. For example, the cited CAIS study concerns children aged 7–17 and does not settle suitability at age six [25]. The cited French PSI-4-SF study is edition-specific and does not establish properties for every language or setting [22]. Early standardisation or internal consistency is not labelled evidence of responsiveness. A study team must still evaluate the exact score, mode, recall period and intended interpretation.

Project and decision representation

The project schema, dipar-project-0.3, holds one or more outcomes. Nine shared study fields describe identity, population, age range, design, intervention, comparison and time origin. Outcome fields record priority, construct, exact instrument, language, evidence, rights, roles, allocation knowledge, calculation, score rules and range, direction, metric, summary, contrast, schedule, assistance, missingness and analysis. There are 37 fields in total; decision records and additional visits are separate structures.

The information source supplies the observations or responses. The subjective rater makes the judgment represented by a response or rating. Calculation describes how the designated total is computed. A parent answering behaviour questions can be both source and rater while software performs arithmetic. A clinician rating an interview can be a separate rater using an allocation-aware parent's information. The schema reserves not-applicable rater awareness for the absence of a subjective rater; automated arithmetic alone does not justify that designation.

Seven decision groups record status, rationale, responsible role or person, and review date. Status is open, decision recorded, or not applicable with an explanation. It is a user declaration, not an assessment of adequacy. A filled evidence field can document an unresolved gap. The planner separately flags missing

required entries, selected incompatible values, invalid numeric limits and inconsistent role-awareness combinations.

Project export generates JSON, outcome specifications in Markdown, an outcome register, definitions, a schedule and a decision record from the same entries. An expected-roster CSV is supplied with headers only: researchers must populate it independently in the approved study system. Imported fields longer than 20,000 characters are rejected explicitly without shortening their content. The web import limit is 10 MB. Older dipar-plan-0.1 records are migrated with the original window text retained and new decisions open; ambiguous legacy scorer information requires review.

Data records and structural checker

One data row represents an expected child–outcome–reporter–occasion combination in a study. Study, child, outcome, reporter and occasion form the assessment key. Family and allocated condition are retained as consistency checks. The status is observed, partial, missing or not applicable. Observed means the designated score is valid under the instrument rules, potentially including allowed item-level missingness. Partial means a response exists but the designated score is unscorable. Non-observed scores remain blank; numeric zero is an observed value when permitted.

The Python checker uses the standard library and reads data, an independent expected roster, an assessment schedule and outcome definitions. It checks headers, keys, expected versus returned assessments, family/arm consistency, finite numeric values, permitted score ranges, role rules and schedule correspondence. Out-of-window timing and departures from planned roles are warnings for investigation, not instructions to remove observations. If reference files are omitted, the report explicitly states which checks were not performed. Supplied but empty reference files fail reconciliation.

The checker does not calculate questionnaire scores, verify source truth, infer reasons for missingness, establish allocation concealment, choose an analysis, impute outcomes or calculate treatment effects. It reports on the structural agreement of the supplied files. An inaccurate roster can therefore produce inaccurate reassurance. The independent preparation and maintenance of that roster are substantive study procedures outside the script.

Fictional teaching project

The teaching project describes a hypothetical programme of six online modules with practitioner calls compared with usual care for one child aged 6–12 per family. It includes a protocol, statistical analysis plan, two outcome specifications, linked reference files, logs and an illustrative methods excerpt. The invented impact and confidence scores have ranges 0–30 and 0–40 respectively. They have no items, norms or measurement evidence and must not be administered to families.

The protocol and analysis plan provide explicit instructional choices, including a hypothetical sample-size calculation, allocation procedure, treatment-policy estimand, primary-window rule, baseline-adjusted model and missingness sensitivity scenarios. These choices are not recommendations for a real trial. In particular, the target difference, variance and missing-data shifts are hypothetical assumptions. The four-child fixture is far too small to implement the proposed full-trial analyses and is used only to demonstrate record checking. No effect estimate is reported.

Word documents provide a blank protocol/analysis companion and the filled teaching narrative. Excel workbooks provide blank record structures and the same filled definitions, schedule, roster, assessments and decision records as the machine-readable example. The formats are intended to improve accessibility for researchers unfamiliar with Markdown or a command line; that usability intention has not been tested.

Internal verification and reproducibility

Twelve JavaScript tests exercise the filled project, JSON round-trip preservation, explicit rejection of oversized fields, shared-field conflicts, duplicate outcome identifiers, role-awareness distinctions, decision-state rules, added occasions, legacy migration, invalid values, CSV quoting and ZIP generation. Fourteen Python tests examine fixture counts and zero preservation, absent and unexpected assessments, duplicates, score range and nonfinite values, non-observed scores, missing reasons, arm and target conflicts, rater awareness, missing reference disclosure, empty references and ZIP readability. These scenario tests were written and executed during the revision; they were not prespecified as an independent study.

The release includes source, fixtures and commands so another person can repeat the checks. A manifest records material sizes and SHA-256 hashes. Site interaction and document/workbook rendering receive internal release inspection, recorded separately from the reproducible code tests. There was no external replication, formal security audit, exhaustive browser compatibility assessment or quantified accessibility certification.

Results

Release components

Version 0.3.0 comprises the project planner, linked editable records, a source-to-prompt map, six measurement cards, the offline checker, the fictional study pack and this development manuscript. The material manifest is the authoritative inventory; a file count should not be interpreted as methodological coverage. Original documentation and templates are licensed CC BY 4.0, and original code is licensed MIT. External sources and instruments retain their own terms.

Component	Result of development	Principal boundary
Project planner	37 fields, shared study details, multiple outcomes and explicit decision states	Completion does not judge scientific adequacy
Linked exports	Definitions, register, schedule, decisions and JSON from one project	The expected roster requires independent preparation
Protocol and delivery records	Prompts for conduct, support, fidelity, changes, analysis and harms	Real governance, permissions and procedures remain study-specific
Measurement cards and map	Six selected evidence anchors and a developer location map	Neither exhaustive nor independently validated
Structural checker	Reconciles references and preserves late records with warnings	Does not verify scoring, source truth or causal assumptions
Teaching study	Two outcomes, 16 records, filled narrative and editable formats	Invented scores and data; no clinical use or inference
Publication and reuse files	Full draft paper, citation exports, licences, archives and manifest	No DOI, peer review or uptake claim is implied

Fictional fixture and executed checks

Across the two baseline assessments per child, all eight scores are observed. At the primary occasion, five scores are observed, two assessments are missing and one response is unscorable. The resulting 16 rows include 13 observed scores, two missing and one partial record. One observed impact score is zero. One confidence observation occurs at week 14, outside the prespecified week 11–13 window, and remains in the data.

The checker reports no structural errors and one outside_window warning for the distributed fixture. Removing a row produces an absent_expected_row error when the independent roster is supplied. Changing an observed score to 99 produces a score_range error. The logs distinguish unknown nonresponse, an incomplete response and the late assessment. Their descriptions are fictional demonstrations, not findings about the causes of missingness in actual research.

All 12 JavaScript and 14 Python scenario tests passed in the recorded run. This supports the inspected behaviours only. The number of tests is not a coverage percentage, and a passing check does not establish a valid study. The source archive permits examination of the assertions and their limits.

Privacy and publication status

The website is a static application with no DiPaR endpoint receiving planner text. Optional browser storage and downloaded backups retain project data on the user's device; ordinary host access logs may still exist. The site instructs users to enter study-level descriptions and keep participant information in approved systems. A local feedback composer prepares a text file without transmitting it. A dedicated public correspondence address requires author confirmation.

The full manuscript is presented consistently as an author-review draft in HTML, Markdown, Word and PDF. It has not been deposited as a preprint, assigned a DOI or peer reviewed. No independent use, citations, downloads or successful implementations are reported. The uptake register begins empty.

Discussion

Contribution and intended use

DiPaR's contribution is a concrete, inspectable arrangement of planning decisions and linked records in one applied setting. Separating informant, rater and calculation exposes a measurement distinction that can be obscured by a single scorer field. Linking outcome definitions and schedules reduces opportunities for manual transcription disagreement. Reconciliation with an independent roster can expose a record that is entirely absent from a returned dataset. These are design rationales supported by the implemented behaviour, not evidence that users make better scientific decisions.

The toolkit complements broader protocol, reporting and measurement resources. The limited comparison with SEPTRE, COBWEB and PROTEUS concerns their stated purposes, not tested feature equivalence or performance. Direct application retrieval for SEPTRE and COBWEB was unavailable during the source check; their functions are described through the official SPIRIT–CONSORT page [12]. No superiority, novelty priority or market-coverage claim is made.

Free templates may lower a practical access barrier, especially when supplied in common document and spreadsheet formats. However, removing a price or registration step does not demonstrate adoption. An open licence also does not transfer third-party instrument rights. A researcher can cite the version and

templates actually used, together with original scientific sources, without being required to cite a future evaluation paper unrelated to that use.

Limitations

The source review was purposive and did not estimate the completeness of the content. The guidance map has not been independently reviewed, and the cards do not establish suitability for every age, form, language or mode. The teaching measures deliberately have no clinical validity. Fully populated instructional records should therefore not be mistaken for a ready-to-run study.

The principal design is narrow. More complex randomisation, multiple children per family, repeated-reporting structures, economic evaluation and jurisdiction-specific governance require additional records and expertise. Data checks operate on selected derived records rather than an entire electronic data-capture history. A consistent set of files may still contain false source values, flawed scoring, unmeasured confounding in non-randomised work or unsuitable analytical assumptions.

Development and verification were conducted with AI assistance under the author's direction. AI source checking and automated tests are not independent scientific review. No end-user tasks, caregiver consultation, psychometric study, reliability rating exercise or comparative evaluation was completed. Security, accessibility and long-term maintenance have not been independently certified. The current website can change; a dated release archive is necessary for reproducible citation, and a persistent repository deposit remains outstanding.

Evaluation agenda

A first external evaluation should examine the source-to-prompt map and filled example with independent trial methodologists, measurement specialists and researchers in digital parenting. Reviewers should record relevance, omissions, misleading wording and applicability, with disagreements and competing interests disclosed. Corrections should be versioned and not used to retrospectively label earlier material validated.

Task-based testing should include beginning and experienced researchers specifying separate parent and child outcomes, identifying an allocation-aware parent source, preserving expected missing rows, interpreting an outside-window warning and exporting linked files. Outcomes could include task completion, errors, time, assistance needed and qualitative feedback. The protocol should define tasks, coding, sample rationale, analysis and any required ethics/consent arrangements before recruitment. No sample or results are asserted here.

A later comparative study could evaluate whether toolkit access improves the adequacy of trial-planning decisions against a prespecified rubric, rated by assessors with documented reliability and appropriate masking. A reduction in blank fields would be an intermediate outcome, not proof of lower risk of bias, better clinical outcomes or more valid studies. Independent uptake should be documented separately from downloads and citations using verifiable public records or permission-based reports.

Conclusions

DiPaR Toolkit 0.3.0 provides linked outcome-planning records, reference-aware structural checks and a reproducible fictional example for digital parent-support research. Internal scenario tests support selected software behaviours. Scientific adequacy, usability, adoption and benefit to research practice remain questions for independent evaluation.

Declarations

Ethics and participation: This descriptive development report used no recruited participants, patient records or participant-level study data. All demonstration records are fictional. Later studies involving people require the relevant local ethics determination and consent arrangements.

Availability: Website: https://diparkit.com/. Versioned materials: https://diparkit.com/materials/dipar-toolkit-v0.3.0.zip. Source and tests: https://diparkit.com/materials/dipar-source-v0.3.0.zip. Each release includes a file manifest. No repository DOI has yet been assigned. Original content is CC BY 4.0; original code is MIT; external sources retain their terms.

Author contribution: Alexander Kuznetsov initiated and directed the resource's scope and development. Final contribution wording, affiliation, correspondence and acceptance of this manuscript require author review before deposit.

AI assistance: Generative AI assisted with drafting, coding, source checking, internal verification and editing. It was not an independent reviewer or author. The named author must verify final claims, references, code and disclosures and take responsibility for a submitted version.

Competing interests: Alexander Kuznetsov created and maintains DiPaR Toolkit and may benefit professionally from its use or citation. Additional relevant relationships require author confirmation before deposit.

Funding: The author must confirm whether development received internal, external or in-kind support. No funding or absence-of-funding claim is inferred here.

Registration: This is a descriptive development report, not a registered trial or prospectively registered evaluation. No registration number is claimed.

References

1. Chan AW, Boutron I, Hopewell S, et al. SPIRIT 2025 statement: updated guideline for protocols of randomised trials. BMJ. 2025;389:e081477. https://doi.org/10.1136/bmj-2024-081477.

2. Butcher NJ, Monsour A, Mew EJ, et al. Guidelines for reporting outcomes in trial protocols: the SPIRIT-Outcomes 2022 extension. JAMA. 2022;328:2345–2356. https://doi.org/10.1001/jama.2022.21243.

3. Baba A, Smith M, Potter BK, et al. SPIRIT-Children and Adolescents 2026 extension statement: enhancing the reporting and usefulness of paediatric randomised trial protocols. BMJ. 2026;392:e085062. https://doi.org/10.1136/bmj-2025-085062.

4. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication checklist and guide. BMJ. 2014;348:g1687. https://doi.org/10.1136/bmj.g1687.

5. COSMIN. Selecting the most suitable outcome measurement instrument. https://www.cosmin.nl/finding-right-tool/select-best-measurement-instrument/. Accessed 23 October 2026.

6. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ. 2025;389:e081123. https://doi.org/10.1136/bmj-2024-081123.

7. Baba A, Smith M, Potter BK, et al. CONSORT-Children and Adolescents 2026 extension statement: enhancing the reporting and impact of paediatric randomised trials. BMJ. 2026;392:e085061. https://doi.org/10.1136/bmj-2025-085061.

8. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898. https://doi.org/10.1136/bmj.l4898.

9. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. JAMA. 2018;319:483–494. https://doi.org/10.1001/jama.2017.21903.

10. Calvert M, King M, Mercieca-Bebber R, et al. SPIRIT-PRO extension explanation and elaboration. BMJ Open. 2021;11:e045105. https://doi.org/10.1136/bmjopen-2020-045105.

11. Eysenbach G; CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res. 2011;13:e126. https://doi.org/10.2196/jmir.1923.

12. SPIRIT–CONSORT. Official website and descriptions of SEPTRE and COBWEB. https://www.consort-spirit.org/. Accessed 23 October 2026.

13. PROTEUS Consortium. SPIRIT-PRO protocol template. 2023. https://theproteusconsortium.org/wp-content/uploads/2023/02/230220-SPIRIT-PRO-PROtocol-template.pdf. Accessed 23 October 2026.

14. Youthinmind. SDQ information and copyright terms. https://www.sdqinfo.org/. Accessed 23 October 2026.

15. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry. 1997;38:581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x.

16. UCLA Child FIRST. RCADS permissions overview. https://rcads.ucla.edu/permissionsoverview. Accessed 23 October 2026.

17. Ebesutani C, Bernstein A, Nakamura BJ, Chorpita BF, Weisz JR, Research Network on Youth Mental Health. A psychometric analysis of the Revised Child Anxiety and Depression Scale—Parent Version in a clinical sample. J Abnorm Child Psychol. 2010;38:249–260. https://doi.org/10.1007/s10802-009-9363-8.

18. PAR. Eyberg Child Behavior Inventory. https://www.parinc.com/products/ECBI. Accessed 23 October 2026.

19. Robinson EA, Eyberg SM, Ross AW. The standardization of an inventory of child conduct problem behaviors. J Clin Child Psychol. 1980;9:22–28. https://doi.org/10.1080/15374418009532938.

20. PAR. Parenting Stress Index Fourth Edition Short Form. https://www.parinc.com/products/PSI-4-SF. Accessed 23 October 2026.

21. Frick PJ. Alabama Parenting Questionnaire: authorised forms and conditions. https://faculty.lsu.edu/pfricklab/alabamaparenting.php. Accessed 23 October 2026.

22. Touchèque M, Etienne A-M, Stassart C, Catale C. Validation of the French version of the Parenting Stress Index–Short Form (Fourth Edition). J Community Psychol. 2016;44:419–425. https://doi.org/10.1002/jcop.21778.

23. Shelton KK, Frick PJ, Wootton J. Assessment of parenting practices in families of elementary school-age children. J Clin Child Psychol. 1996;25:317–329. https://doi.org/10.1207/s15374424jccp2503_8.

24. Essau CA, Sasagawa S, Frick PJ. Psychometric properties of the Alabama Parenting Questionnaire. J Child Fam Stud. 2006;15:597–616. https://doi.org/10.1007/s10826-006-9036-y.

25. Langley AK, Falk A, Peris T, et al. The Child Anxiety Impact Scale: examining parent- and child-reported impairment in child anxiety disorders. J Clin Child Adolesc Psychol. 2014;43:579–591. https://doi.org/10.1080/15374416.2013.817311.

Appendix A. Reproducing the structural example

Unzip the toolkit and change into the materials directory. Use Python 3; no external packages are required for the checker.

```
python3 validate_data.py synthetic-data.csv --roster synthetic-expected-roster.csv --
schedule synthetic-measurement-schedule.csv --definitions synthetic-outcome-
definitions.csv
```

Expected result: 16 rows; observed 13, missing 2, partial 1; no errors; one outside_window warning. The script exits successfully when there are no errors, even if warnings require review. It never edits the input files.

Appendix B. Repeating the software scenarios

Unzip the source archive, enter its root directory, and use Node.js and Python 3. The test code uses standard runtime libraries.

```
node --test tests/planner.test.mjs
```

```
python3 -m unittest discover -s tests -p 'test_*.py'
```

Run JavaScript first: its ZIP test creates an archive subsequently checked by Python. Scenario assertions, rather than only aggregate pass counts, are available for inspection. Document/workbook generation has additional dependencies described in the source README and is separate from these tests.