PharmaSUG China 2026 - Paper DS-214

When Data Capture Becomes Data: e-Diary Metadata, Denominators, and ADaM Traceability in Vaccine Studies

Jialiang LIU and Lin CHEN, Sanofi

ABSTRACT

Electronic clinical outcome assessment (eCOA) systems, including electronic patient-reported outcome (ePRO) e-Diaries, are increasingly used to collect vaccine solicited reactions. Unlike paper diaries, they can provide trusted entry time, upload or receipt time, collection mode, device context, and audit metadata. These elements may arrive through different extracts and at different grains. The analysis- and traceability-relevant subset should therefore be retained without reproducing the full audit trail in submission datasets. This paper focuses on four consequences: distinguishing a true negative from a missing assessment; separating clinical, intended-day, entry, and receipt times; preserving subject and investigator records; and deriving question-specific denominators. A five-subject Days 1–8 example separates expected diary participation (N_DIARY), endpoint-specific evaluability (M), electronic submission, and cumulative timing windows (≤24, ≤48, and ≤72 hours). It also shows how paper fallback can contribute clinical evidence without electronic timing. We propose an expected-assessment staging frame, selected SDTM metadata, explicit ADaM flags, and reviewer-facing shells for participation, timing, and maximum severity. The examples are constructed; timing-restricted views are supportive or sensitivity analyses unless otherwise pre-specified.

Keywords: eCOA, ePRO, e-Diary, Vaccine Reactogenicity, SDTM, ADaM

1. INTRODUCTION: VACCINE STUDIES DEPEND ON SUBJECT-REPORTED DATA

Vaccine trials routinely collect solicited injection-site reactions, systemic symptoms, and temperature during a fixed period after vaccination, often 7 or 14 days. These data are recorded directly by the subject, historically on paper and now commonly through ePRO e-Diaries. The electronic instrument does more than replace the paper form: each assessment may be linked to a system timestamp, source key, collection mode, device context, and audit metadata. Those elements can affect whether a result is observed, when it becomes available, and which denominator it supports.

eCOA is the electronic implementation of a clinical outcome assessment and is classified by the assessor: ePRO is reported by the participant, eObsRO by an observer, eClinRO by a clinician, and ePerfO through a performance task [1]. An adult vaccine e-Diary is usually ePRO; a caregiver diary may be eObsRO. Digital health technologies are broader and are not automatically eCOA [2]. The same capture-metadata issues arise in other therapeutic areas with frequent symptom, function, seizure, pain, or caregiver-reported assessments.

This paper asks how e-Diary capture context affects analysis. It retains the TAUG-VAX CE/FACE/VS clinical model while linking selected metadata through SUPP-- or another documented structure and carrying the analysis-relevant subset into ADaM. The main proposition is that diary participation, endpoint evaluability, collection mode, and entry timing answer different questions and require different denominators.

Figure 1 summarizes the path from source capture through staging, SDTM, ADaM, and reporting. Sections 2–3 define the data and workflow; Section 4 develops the analysis and table shells; Sections 5–7 address traceability and implementation; Section 8 brings the concepts together in one five-subject example.

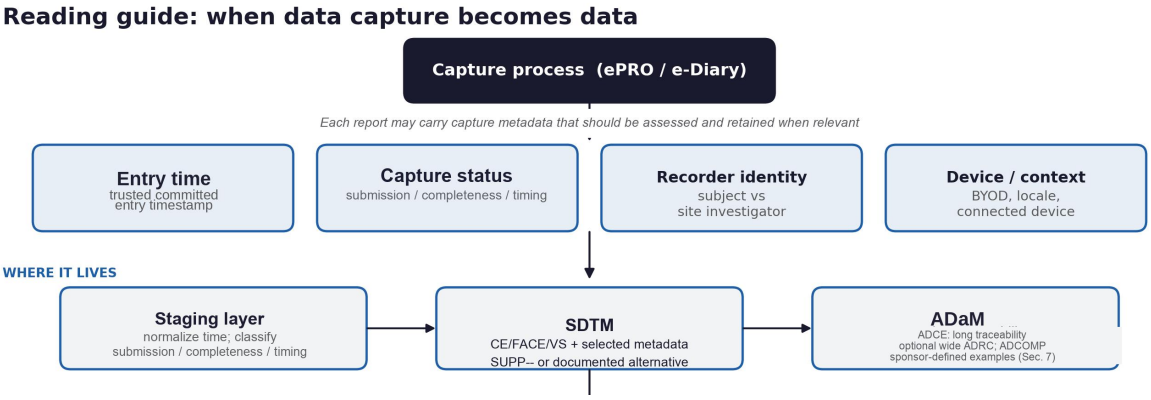

Figure 1. Reading guide from e-Diary capture metadata to analysis and reporting.

2. LIMITATIONS OF CURRENT VACCINE STANDARDS

2.1 THE VISIT-BASED, FLAT MODEL MEETS TIME-DENSE DATA

An e-Diary produces a high-frequency, time-stamped sequence rather than a set of periodic site observations. One subject may generate several daily results across multiple solicited reactions, which complicates nominal-day alignment and cross-domain traceability.

The TAUG-VAX flat model commonly used for OVRR submissions [3–6] places one summary record per subject, reaction, and dose context in CE; daily reaction assessments in FACE; and daily temperatures in VS. Planned negative records are retained so that completed negatives can be distinguished from missing assessments. The model may be adapted for large trials in discussion with the review team.

- **CE (Clinical Events) carries the event summary:** CEOCCUR = "Y" if the solicited reaction was reported on any day, "N" if every day was completed and negative.
- **FACE carries the daily records:** the objective injection-site diameter and the subjective daily severity grades.
- **VS carries the daily temperatures,** kept separate from FACE.

Solicited reactions use CECAT = "REACTOGENICITY" and CESCAT = "ADMINISTRATION SITE" or "SYSTEMIC". RELREC links CE summaries to daily FACE/VS records and to any AE required under protocol-defined escalation criteria.

2.2 ABSENCE IS DATA, BUT MISSINGNESS IS NOT ABSENCE

A completed negative is evidence; a missed assessment is not. OVRR therefore uses --STAT = "NOT DONE" with --REASND for a missed assessment. CEOCCUR is "N" only when all required daily assessments are complete and negative; it remains null when the series is incomplete and no positive reaction has been observed. e-Diary process metadata helps preserve this distinction.

2.3 ONE RESULT, SEVERAL CLOCKS, AND AN APPLICABLE ANCHOR

One result can carry several clocks. Vaccination datetime anchors Day 1; clinical time describes onset or resolution; assessment datetime describes the daily evaluation when captured; trusted entry time records committed entry; and upload or receipt time documents transport. ENTRYHRS is elapsed time from the applicable intended-day anchor to trusted entry. For Day 1 the anchor is vaccination datetime; later anchors follow the planned schedule. Receipt time should not replace entry time for timing analysis.

Timing may be available at form or item level. Day-1 injection-site, systemic, and temperature forms can therefore have different entry times even when they share an intended day. Under the illustrative right-closed rule used here, 0 ≤ ENTRYHRS ≤ 72 is within the timing window; ENTRYHRS > 72 is submitted after the window. ENTRDTC, the anchor, and ENTRYHRS should be retained at the finest supported grain.

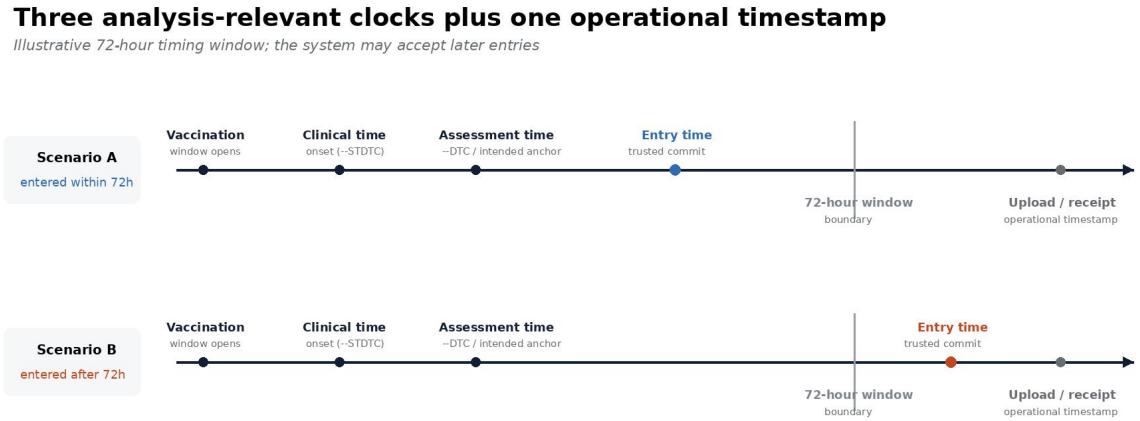

Figure 2. Alternative entry paths for one Day-1 form. Trusted entry, not later upload or receipt, determines membership in the illustrative right-closed 72-hour window.

SDTM supports complete or partial ISO 8601 date/time values and time-zone offsets [7], but --DTC alone does not distinguish clinical, assessment, entry, and receipt clocks. It also does not provide a standard timing-window classification. These distinctions therefore require documented source fields, sponsor metadata, or another tabulation structure.

2.4 COMPLIANCE METADATA HAS NO CANONICAL HOME IN SDTM

Paper completion does not prove contemporaneous entry [8]. ePRO makes form-level timing observable, but SDTM has no canonical variable for the resulting classification. Sponsors may use selected SUPPFACE/SUPPVS qualifiers or another documented structure; the implementation and timestamp definitions must be explained in metadata and reviewer guides.

2.5 THE DUAL-RECORDER PATTERN

When site-investigator confirmation is collected, both the subject and investigator records should remain in FACE and be distinguished by FAEVAL. A pre-specified ADaM precedence rule may select the investigator value when usable; the subject's trusted entry time remains the timing source and the investigator review time is retained separately [9].

USUBJID	FAOBJ	FATESTCD	FAEVAL	FATPTNUM	FAORRES
S-001	PAIN	SEV	STUDY SUBJECT	3	GRADE 2
S-001	PAIN	SEV	INVESTIGATOR	3	GRADE 3
S-001	TENDERNESS	SEV	STUDY SUBJECT	3	GRADE 1

Table 1. Subject and investigator records retained in FACE. ADaM flags implement analysis precedence without overwriting the source-reported value [10].

2.6 OTHER COMPLICATIONS: COLLECTION MODE, MULTI-DOSE, AND TIMING-WINDOW RULES

Collection mode can affect observability and should remain explicit when paper and electronic data are pooled. A difference between studies or treatment arms may partly reflect capture mode rather than tolerability.

In multi-dose studies, diary participation and timing should be summarized separately by vaccination because the pattern may differ across doses or treatment arms.

The protocol, eCOA specification, SAP, or system configuration should define the intended-day anchor, timing window, and handling of records accepted after the boundary. The unrestricted primary reactogenicity analysis in this paper includes all valid, evaluable submitted results. Views restricted to ≤24, ≤48, or ≤72 hours are supportive or sensitivity analyses and must re-derive their own n and M.

3. DATA FLOW AND PROGRAMMING REQUIREMENTS

3.1 DELIVERY PATH AND EXTRACT COMPLETENESS

e-Diary data may be natively integrated with EDC, transferred through an interface, or delivered from a standalone platform. The architecture matters less than the actual extract: clinical values and process metadata may arrive at different grains, through different reports or application programming interfaces (APIs), and on different schedules. Integration therefore does not prove that entry time, receipt time, engagement evidence, device context, or audit metadata are present in the routine programming feed.

Before collection begins, the Data Transfer Specification should identify each required field, grain, trusted timestamp, source key, extraction path, and delivery cadence. Study-defined engagement evidence—such as activation, first login, or form opening—is also needed to distinguish an engaged-but-empty subject from one who never engaged.

Reminder history and the full audit trail may remain in the validated source platform; only analysis- or traceability-relevant attributes need to enter the programming flow.

3.2 A SPONSOR-CONTROLLED EPRO STAGING LAYER

A sponsor-controlled staging layer creates a common expected-assessment frame before SDTM. It normalizes timestamps, preserves source keys, links clinical results to forms or items, and derives submission, completeness, collection mode, and timing attributes. In an integrated platform this may be a logical transformation rather than a separate database.

Keep these attributes separate. SUBMSTAT distinguishes SUBMITTED from NOT SUBMITTED for each expected unit; FORMSTAT classifies submitted content as COMPLETE, PARTIAL, or NOT EVALUABLE; and TMEVALFL identifies submitted electronic units with a valid anchor and trusted timestamp. ENTRYHRS is elapsed time from the applicable anchor. Under the illustrative rule, WITHIN WINDOW means 0 ≤ ENTRYHRS ≤ 72 and AFTER WINDOW means ENTRYHRS > 72; cumulative WIN24FL, WIN48FL, and WIN72FL are right-closed. The configured system prevents entry before the applicable anchor, so early entry is not a routine reporting category. Timing variables do not by themselves determine primary-analysis inclusion.

Grain must be explicit. Submission and completeness are usually form- or card-level; clinical results, endpoint evaluability, and M are reaction- or item-level. Use item-level entry time when available; otherwise apply a documented parent-form inheritance rule.

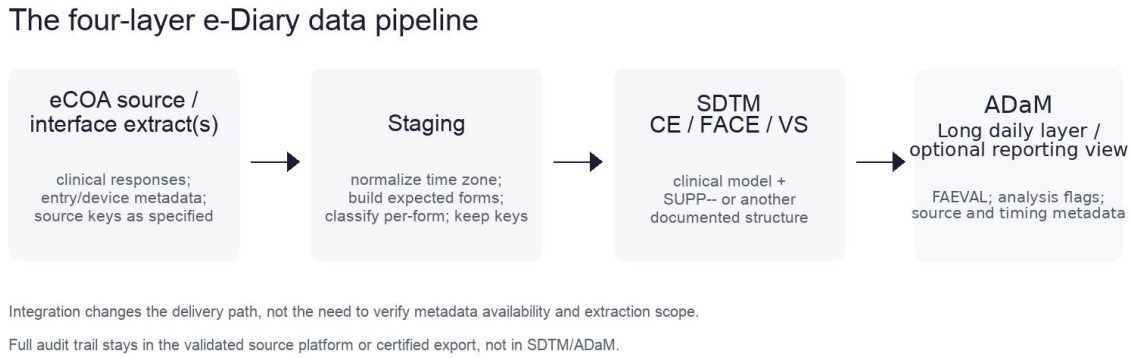

Figure 3. Source or interface data are normalized in staging before selected clinical and capture metadata flow to SDTM, ADaM, and reporting.

3.3 RECONCILIATION AND DEVICE CONSIDERATIONS

Reconciliation remains necessary across the e-Diary, investigator confirmation, AE/SAE records, and related medication data. Checks should reflect protocol escalation rules and preserve links rather than rely only on equality joins by study day [11,12].

Device type and source precedence should be retained when relevant. Whether a subject-confirmed value or a validated device-transmitted value is analyzed depends on the protocol and must be documented.

3.4 SUPP-- CREATION AND VALIDATION

After expected units are established, selected metadata may be attached to FACE/VS through SUPP-- or another documented structure. Expected but unsubmitted assessments first require a parent record with --STAT = "NOT DONE" and --REASND. Form-level attributes should not be copied to every item without a documented inheritance rule. Sponsor validation must also check issues that conformance tools cannot detect, such as an invalid anchor or CEOCCUR = "N" despite incomplete daily data. Open-source tools such as admiral and xportr can support derivation and transport [13,14].

4. EFFECTS ON ANALYSIS AND ENDPOINTS

4.1 NEW AND NEWLY EXPECTED ANALYSES

Timestamped data support sub-day onset and duration, reaction-medication alignment, and cumulative timing views. These analyses require separated clocks, consistent time-zone handling, and analysis-level access to entry time and the applicable anchor.

4.2 COMPLIANCE AS AN ANALYSIS OBJECT

Diary metrics answer different questions. Table 2 summarizes whether any diary data were provided through any protocol-permitted mode and whether all required diary data were complete, by day and over the full period. Any e-Diary data is a separate electronic-use metric and should be labeled explicitly. Tables 3–4 apply cumulative timing windows to endpoint-specific incidence and maximum severity. Table 5 separates submission, completeness, and timing at their stated grains; Table 6 compares these patterns across treatment groups.

Compliance metric (per arm and vaccination)	Day 1	...	Day n	Full period
N_DIARY: subjects expected to complete the diary	n	…	n	n
N (%) providing ANY diary data (all protocol-permitted modes)	n (%)	...	n (%)	n (%)
N (%) providing COMPLETE diary data (all protocol-permitted modes)	n (%)	...	n (%)	n (%)

Table 2. Diary participation and completeness by arm and vaccination. Day-level and full-period definitions are separate and include protocol-permitted collection modes.

Table 3 is an illustrative reviewer-facing shell, not a regulatory requirement. For Day d, the intended-day anchor is vaccination datetime + 24 × (d − 1) hours. The right-closed windows are 0 ≤ ENTRYHRS ≤ 24, ≤48, and ≤72. Individual reactions use their own trusted entry time. For Any injection-site reaction, clinical composite status and timing evaluability are derived separately: a positive component establishes a clinical positive, but timing is evaluable only when all required component times are evaluable, using the latest component time.

Solicited reaction	Intended diary day	Reporting window	Vaccine A (N_SAFETY=###)		Comparator (N_SAFETY=###)	
			n/M	%	n/M	%
Any injection-site reaction	Day 1	≤24 hours	###/###	##.#	###/###	##.#
		≤48 hours	###/###	##.#	###/###	##.#
		≤72 hours	###/###	##.#	###/###	##.#
	Day 2	≤24 hours	###/###	##.#	###/###	##.#
		≤48 hours	###/###	##.#	###/###	##.#
		≤72 hours	###/###	##.#	###/###	##.#
	...	...	...	...	...	...
	Day 8	≤24 hours	###/###	##.#	###/###	##.#
		≤48 hours	###/###	##.#	###/###	##.#
		≤72 hours	###/###	##.#	###/###	##.#
Tenderness	Day 1	≤24 hours	###/###	##.#	###/###	##.#
		≤48 hours	###/###	##.#	###/###	##.#
		≤72 hours	###/###	##.#	###/###	##.#
Other selected reactions	...	...	...	...	...	...

Table 3. Selected solicited injection-site reactions by intended diary day and cumulative right-closed reporting window. n and M are derived under the same endpoint, component, and timing rule.

Table 4 applies the same windows before deriving maximum severity across Days 1–8. Individual reactions use eligible daily records; the composite first derives daily occurrence and timing eligibility. The unrestricted primary summary remains separate from these supportive or sensitivity views.

		GROUP A		GROUP B	
Dose	**Local reaction**	**n/M**	**% (95% CI)**	**n/M**	**% (95% CI)**
Dose 1	Any injection-site reaction	###/###	##.# (##.#; ##.#)	###/###	##.# (##.#; ##.#)
	Tenderness				
	None	###/###	##.# (##.#; ##.#)	###/###	##.# (##.#; ##.#)
	Any	###/###	##.# (##.#; ##.#)	###/###	##.# (##.#; ##.#)
	Mild	###/###	##.# (##.#; ##.#)	###/###	##.# (##.#; ##.#)
	Moderate	###/###	##.# (##.#; ##.#)	###/###	##.# (##.#; ##.#)
	Severe	###/###	##.# (##.#; ##.#)	###/###	##.# (##.#; ##.#)
...	Other selected reactions/doses	...	...	...	...

Table 4. Illustrative ≤24-hour maximum-severity shell for Days 1–8 (parallel: ≤48 and ≤72 hours). Composite and individual-reaction M values are derived under the same window-specific rules as their numerators.

Table 5 keeps three expected-unit attributes separate: submission, content completeness, and electronic timing. Its cumulative ≤24-, ≤48-, and ≤72-hour columns overlap; >72 hours identifies submitted after-window units.

Panel A. Submission and form completeness						
Form family / intended day	**Expected**	**Submitted**	**MISSED**	**Complete**	**Partial**	**Form Not Evaluable**
Injection-site card / Day 1	n	n (%)	n (%)	n (%)	n (%)	n (%)
Injection-site card / Day 2	n	n (%)	n (%)	n (%)	n (%)	n (%)
Systemic card / Day 1	n	n (%)	n (%)	n (%)	n (%)	n (%)
Temperature card / Day 1	n	n (%)	n (%)	n (%)	n (%)	n (%)
...	...	...	...	...	...	...
Panel B. Cumulative electronic entry timing over expected electronic assessment units at the displayed grain						
Electronic assessment unit / intended day	**Timing Evaluable**	**≤24h**	**≤48h**	**≤72h**	**>72h**	**Timing Not Evaluable**
Injection-site / pain / Day 1	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Injection-site / pain / Day 2	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Systemic / each reaction / Day 1	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Temperature / temperature / Day 1	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
...	...	...	...	...	...	...

Table 5. Submission and completeness are form/card-level; timing may be form- or item-level according to the Data Transfer Specification. Cumulative timing columns use the applicable intended-day anchor and trusted timestamp. Timing Not Evaluable denotes a submitted electronic unit without a valid anchor or timestamp; MISSED denotes an expected unit not submitted by the analysis cutoff. Paper contributes to Panel A only.

Table 6 compares full-period participation, completeness, and cumulative electronic timing across treatment groups and vaccinations.

Compliance metric	**Group A**	**Group B**	**Group C**	**Total**
N_DIARY: subjects expected to complete the diary	n	n	n	n

Compliance metric	Group A	Group B	Group C	Total
Subjects providing ≥1 diary record through any protocol-permitted mode during the solicited period	n (%)	n (%)	n (%)	n (%)
Subjects with all required diary records complete throughout the solicited period	n (%)	n (%)	n (%)	n (%)
e-Diary forms entered ≤24 hours among expected e-Diary forms	n (%)	n (%)	n (%)	n (%)
e-Diary forms entered ≤48 hours among expected e-Diary forms	n (%)	n (%)	n (%)	n (%)
e-Diary forms entered ≤72 hours among expected e-Diary forms	n (%)	n (%)	n (%)	n (%)
e-Diary forms entered >72 hours among expected e-Diary forms	n (%)	n (%)	n (%)	n (%)

Table 6. Cross-arm comparison. Full-period metrics include protocol-permitted modes; electronic timing uses the pre-specified expected electronic frame. The SAP should state whether that frame remains fixed after an approved switch to paper.

Together, Tables 2–6 move from expected units to subject participation and endpoint-specific results. All should be reproducible from the same expected-assessment frame, linked clinical records, collection mode, and trusted timing metadata.

This hierarchy matters because unequal missingness or delayed availability across arms can change the observed reaction profile. A wider timing window can also appear stable while an earlier view rests on a smaller M.

For subjects scheduled to complete a diary, assign four hierarchical states per vaccination: A, no engagement; B, paper fallback; C, at least one submitted e-Diary form; and D, electronic engagement but no electronic or paper symptom data. Define engagement in advance and keep per-form collection mode separate.

Diary state	Electronic engagement evidence?	≥1 e-Diary form?	In expected-to-diary population?	Any data?	Timeliness evaluable?
A - No engagement	No	No	Yes	No	No
B - Paper fallback	May be present	No (e-Diary)	Yes	Yes (paper)	No
C - e-Diary engaged	Yes	Yes	Yes	Yes	Possible; determined at form level by TMEVALFL
D - Engaged, empty	Yes	No	Yes	No	No

Table 7. Subject-by-vaccination diary states. All four belong to N_DIARY; only endpoint-level evidence determines whether a subject contributes to M.

4.3 THE M DENOMINATOR VERSUS DIARY-COMPLIANCE DENOMINATORS

Vaccine tables distinguish N_SAFETY, N_DIARY, and M. N_DIARY contains subjects scheduled for diary collection; M contains subjects with evaluable data for the specified endpoint, dose, and analysis window [4]. Missing solicited assessments are not automatically non-events.

States B and C can contribute to M when they contain an evaluable result. States A and D cannot contribute from diary data unless another permitted source supplies the endpoint. Activation or form assignment alone is not evidence of presence or absence.

There is no universal compliance denominator. Participation may use N_DIARY; form metrics may use expected electronic units; timing distributions may use timing-evaluable submissions; and reactogenicity uses endpoint-specific M. Each denominator must be named and reproducible.

For a composite, any positive required component establishes a positive result. A negative requires all required components to be observed and negative; no positive plus a missing component is not evaluable. Timing evaluability is separate and, under the conservative rule, requires all component times and uses the latest one.

Cumulative timing may be reported over expected electronic units or over timing-evaluable submissions, depending on the question. A submitted unit with ENTRYHRS > 72 is AFTER WINDOW; a unit never submitted by cutoff is MISSED; paper has no electronic timing. The SAP should define the anchor, threshold, grain, cutoff, denominator, and paper handling.

Every timing-restricted endpoint view must re-derive both n and M within the selected window. The same numeric threshold does not make a form-level timing rate and an endpoint incidence interchangeable.

The key takeaway is that denominator logic must be derived from expected forms and evaluable records, not from activation or submitted records alone.

Exhibit 1. Five-profile running example

The five constructed profiles used throughout the paper have 24 expected card-family units per subject (8 days × 3 families), plus a larger set of reaction-item assessments.

Subject	Full-period evidence	Day-1 pain	Day-1 pain electronic timing
S-001	24/24 e-Diary card-family assessments complete	Grade 1, e-Diary	≤24-hour view
S-002	24/24 e-Diary card-family assessments complete	Grade 2, e-Diary	ENTRYHRS = 29; in ≤48 h and ≤72 h
S-003	21/24 e-Diary assessments; Days 2–8 only	Missing	No Day-1 pain timing result
S-004	24/24 paper card-family assessments complete	Grade 0, paper	Electronic timing not applicable
S-005	Engaged but empty	Missing	No submitted assessment

At the subject-period level, 4/5 provide any diary data and 3/5 are complete. For unrestricted Day-1 pain, n = 2 and M = 3. Paper supplies clinical evidence but not electronic timing; each restricted view must re-derive n and M.

4.4 COMPLIANCE AS A REACTOGENICITY ASCERTAINMENT MECHANISM

Diary participation and entry timing are part of reactogenicity ascertainment because short-lived symptoms may be reported retrospectively. The unrestricted primary summary retains valid, evaluable submitted results; supportive or sensitivity views can compare cumulative ≤24-, ≤48-, and ≤72-hour results, after-window records, and completeness patterns.

Analysis	Purpose	Key data elements
Form-day completion rate	Show whether subjects had enough observable diary time.	Expected/completed/missing forms, --TPTNUM or harmonized planned-day variable, window metadata
Solicited-reaction incidence by entry timing	Evaluate whether rates change when cumulative timing restrictions are applied or after-window records are assessed separately.	Occurrence, ENTRDTC/ENTRYHRS, timing flags, TMRSTFL, population flag

Analysis	Purpose	Key data elements
Onset/duration sensitivity	Test whether onset and resolution depend on the timing of entry, including after-window records.	--STDTC/--ENDTC, --DTC, ENTRDTC, ENTRYHRS, TMRSTFL
Subject-investigator concordance	Quantify subject vs site-investigator severity.	FAEVAL, PRIMRFL/SUBJFL, modification flag/reason
Cumulative timing-window sensitivity	Show how incidence or maximum severity changes as additional evaluable data become available at pre-specified thresholds.	Output-specific anchor, ENTRDTC, timing flags, window-specific M
Nested subject-level e-Diary populations	Separate vaccinated, any-use, complete, and fully on-time operational populations without replacing endpoint-specific M.	ANL01FL-ANL04FL, diary state, expected-assessment frame

Table 8. Analyses enabled by capture metadata. These complement, rather than replace, standard safety summaries.

4.5 MISSING DATA, INTERCURRENT EVENTS, AND BORROWED METHODS

Rescue medication can be an intercurrent event when it changes the value or interpretation of a reaction; the SAP must state the strategy. Delayed or missed entry and device failure are usually capture or missing-data issues, although the clinical event causing missingness may itself be an intercurrent event [15].

A result may be absent, partial, entered after the window, or missing after withdrawal. A later report may still support severity while onset and duration remain more sensitive to recall. Retain the capture category rather than forcing every case into a binary missing/not-missing label.

Other therapeutic areas provide useful missing-data and range-alignment methods [11,12,16,17], but vaccine programs still need vaccination-specific windows, solicited-reaction rules, and recorder precedence.

5. TRACEABILITY AND REGULATORY DOCUMENTATION

ICH E6(R3) treats relevant metadata as part of source records and requires traceable initial entry and changes [18]. The full audit trail remains in the validated source system [19]; submission data carry only what is needed for analysis and traceability, such as trusted entry time, timing status, modification indicators, and source keys.

The cSDRG should define the source platform and extract, trusted entry time, time-zone policy, multiple-entry rule, and timing classifications. The ADRG should document recorder precedence, analysis windows, denominator derivations, and partial-datetime handling.

FDA, EMA, NMPA, and ICH materials support reliable timestamps, auditability, and documented oversight [18,20–24], but do not prescribe one universal e-Diary mapping. The source-to-output chain must therefore be fit for purpose and explicitly documented.

6. IMPLEMENTATION RECOMMENDATIONS

The practical recommendations have two levels: study-level implementation and sponsor-level vocabulary. A short standards outlook follows.

6.1 STUDY-LEVEL ACTIONS

Retain full datetimes and source keys; distinguish clinical, assessment, entry, receipt, and modification times; keep missingness separate from a confirmed negative; parameterize anchors and timing thresholds; define M for every endpoint and window; and retain both subject-level diary state and per-form collection mode.

6.2 SPONSOR-LEVEL VOCABULARY

Table 9 proposes a concise sponsor vocabulary for capture metadata when available and relevant.

QNAM	Type	QLABEL	Example / Notes
ENTRDTC	char (ISO 8601)	Entry Datetime	Trusted committed entry; not upload/receipt
EDIARYID	char	e-Diary Source Record Identifier	Traceability key; one per source entry
SUBMSTAT	char (CT)	Form Submission Status	Expected: SUBMITTED / NOT SUBMITTED; latter = MISSED
FORMSTAT	char (CT)	Form Completeness Status	Submitted only: COMPLETE / PARTIAL / NOT EVALUABLE
TMEVALFL	char (Y/N)	Electronic Timing Evaluable Flag	Submitted e-Diary only: Y when anchor and valid trusted timestamp permit ENTRYHRS derivation; N when timing cannot be evaluated; retain reason.
COMPSTAT	char (CT)	Entry Timing Status	TMEVALFL = Y only; example categories: WITHIN WINDOW / AFTER WINDOW. The boundary and anchor are output-specific.
COMPFL	char (Y/null)	Compliant Entry Flag	Y when the timing-evaluable entry is within the documented window; null for after-window entries. Primary-analysis inclusion is controlled separately.
ENTRYHRS / AFTERMIN	char (numeric text in QVAL)	Elapsed Entry Hours / Minutes After Window	ENTRYHRS is elapsed time from the applicable anchor to entry; AFTERMIN is optional for records entered after the timing-window boundary. Numeric in ADaM.
WINOPDTC / WINCLDTC	char (ISO 8601)	Window Open / Timing-Window Close Datetime	Retain the applicable anchor and timing-window boundary, or the parameter needed to reproduce them.
UPLDDTC / RCPTDTC	char (ISO 8601)	Upload / Receipt Datetime	Operational metadata; distinct from ENTRDTC
DEVTYPE / LOCALE	char (CT)	Device Type / Entry Locale	Retain when needed for device/locale analysis
WIN24FL / WIN48FL / WIN72FL	char (Y/null)	Cumulative Timing Flags	Y when 0 ≤ ENTRYHRS ≤ 24, ≤48, or ≤72, respectively; null otherwise. The thresholds are cumulative and right-closed.
ENTRBAND	char (CT)	Entry Timing Band	Trusted ENTRDTC relative to the applicable anchor: 0-24H; >24-48H; >48-72H; >72H. The first three bands are right-closed.

Table 9. Illustrative sponsor vocabulary. SUPP-- is not mandatory; analysis-critical, high-volume, or poorly parented metadata may require another documented structure [25].

6.3 BRIEF STANDARDS OUTLOOK

Future guidance could standardize entry time, timing evaluability, mixed-mode collection, and expected-form linkage. Dataset-JSON modernizes transport [26,27] but does not determine where capture metadata belongs; current studies still need a documented sponsor implementation.

7. ADAM IMPLEMENTATION FOR TRACEABLE ANALYSIS

The following design is illustrative; the required feature is a traceable long daily grain, not a particular dataset name.

7.1 A LONG DAILY ANALYSIS LAYER AND AN OPTIONAL REPORTING VIEW

A long daily layer keeps one row per subject, vaccination, reaction, parameter, planned day, and recorder, with location or laterality when relevant. This paper uses ADCE as an illustrative sponsor-defined label and ADRC for an optional reporting view; neither is canonical ADaMIG [28,29]. Timing and recorder data remain day-specific, and ENTRDTC becomes numeric ENTRDTM in ADaM.

7.2 KEEP BOTH RECORDS, FLAG THE ANALYZED ONE

Keep subject and investigator records and select the analysis record with explicit flags:

- **PRIMRFL selects the pre-specified primary record within subject, vaccination, reaction, parameter, day, and relevant location. When the investigator record is selected, timing remains linked to the subject entry.**
- **SUBJFL identifies the evaluable subject-only record for recorder sensitivity analyses.**
- **TMRSTFL identifies the selected record meeting an output-specific timing rule. For composites, derive timing eligibility after component aggregation rather than from component flags alone.**

A timing-restricted result applies the record-selection and timing flags together. Clear descriptive names avoid conflict with subject-level operational flags; both source rows remain traceable.

7.3 WHERE SHOULD "WHO RECORDED THIS" LIVE IN ADAM?

FAEVAL should be carried forward. RECSRC and documented selection flags can supplement it; DTYPE is shown in Table 10 only as a less-preferred sponsor convention.

Option	Mechanism	Strengths	Weaknesses
Extend DTYPE under a documented sponsor convention	DTYPE = SUBJECT/INVESTIGATOR as sponsor-defined extensible values	Uses an existing variable; self-documenting	Stretches DTYPE beyond design intent [30]
Carry FAEVAL / add RECSRC	Carry FAEVAL forward; RECSRC for cross-domain ADCOMP	Preserves SDTM lineage; covers subject/investigator/device/site	Needs metadata documentation for RECSRC
Descriptive record-selection flags	PRIMRFL primary, SUBJFL subject-only, TMRSTFL timing-restricted	Pre-encodes recorder and timing views without reusing the subject-level ANL01FL-ANL04FL sequence	Requires sponsor metadata and does not by itself preserve per-row recorder identity

Table 10. Options for representing recorder identity and selection in ADaM.

7.4 CARRYING METADATA FORWARD

Carry into ADaM the metadata needed to reproduce timing, missingness, collection mode, recorder selection, and denominators. Keep subject-level diary state separate from per-form attributes unless a clearly parameterized sponsor dataset is used.

Metadata	SDTM location	Into ADaM?	Used for
Clinical onset/resolution	--STDTC / --ENDTC	Yes (datetime)	Duration, onset
Assessment datetime / intended anchor	--DTC; planned day/timepoint variables	Yes (ADT/ATM and planned anchor variables)	Visit/day mapping, window anchoring
Entry datetime (trusted entry)	SUPP-- ENTRDTC	Yes	Timing compliance, reporting-window sensitivity, traceability
Upload / receipt time	SUPP-- UPLDDTC/RCPTDTC	Optional	Audit, data flow
Electronic timing evaluability and status	SUPP-- TMEVALFL / COMPSTAT; optional ENTRBAND, or documented alternative	Yes (+ COMPFL)	Timing-restricted analysis; compliance endpoint when pre-specified
Submission / form completeness	--STAT/--REASND; optional SUBMSTAT/FORMSTAT or documented alternative	Yes	Expected forms, completeness, missingness
Recorder identity	FAEVAL	Yes (FAEVAL; RECSRC)	Dual-recorder precedence
Diary state (A-D)	Staging; optional SUPP-- on a documented parent record	Yes	Denominator definition
Source/interface record key	SUPP-- EDIARYID	Yes	Traceability
Audit trail (full)	Not in SDTM	No	Inspection, audit-trail review

Table 11. Illustrative location and analysis use of e-Diary metadata. The full audit trail remains in the validated source platform.

7.5 NESTED SUBJECT-LEVEL POPULATIONS FOR OPERATIONAL E-DIARY VIEWS

For operational views, a sponsor may derive nested subject-by-vaccination flags. These do not replace endpoint-specific M or the record-level flags above.

- ANL01FL - the study-defined vaccinated or safety-analysis population eligible for diary assessment.
- ANL02FL - at least one submitted e-Diary assessment; a subset of ANL01FL.
- ANL03FL - all required e-Diary assessments complete under the pre-specified expected-assessment frame; a subset of ANL02FL.
- ANL04FL - all required electronic assessments meet the pre-specified on-time rule; a subset of ANL03FL.

ANL04FL $\subset$ ANL03FL $\subset$ ANL02FL $\subset$ ANL01FL. A paper-complete subject can be complete overall without qualifying for complete or fully on-time e-Diary populations.

8. FULL WORKED EXAMPLE: FIVE PROFILES, MULTIPLE GRAINS, AND QUESTION-SPECIFIC DENOMINATORS

8.1 FULL-PERIOD DIARY PARTICIPATION

Five constructed subjects are scheduled for Days 1–8 collection. Each day has injection-site, systemic, and temperature cards, yielding 24 expected card-family units per subject. S-001 and S-002 submit all 24 electronically; S-002's Day-1 pain has ENTRYHRS = 29. S-003 submits Days 2–8 only. S-004 completes all 24 on paper after device failure. S-005 shows electronic engagement but supplies no electronic or paper symptom data by the final cutoff.

Card-family completeness is not endpoint evaluability. Each card contains several reaction items, and every item has its own Days 1–8 trajectory and potentially its own timing.

USUBJID	Diary state	e-Diary units	Paper units	Any data during period?	Complete over period?	Days 1–8 pain entries ≤24 h / timing-evaluable
S-001	C - engaged	24	0	Yes	Yes	8/8
S-002	C - engaged	24	0	Yes	Yes	4/7 (1 timing NE)
S-003	C - engaged	21 (Days 2–8 only)	0	Yes	No	4/7 (Day 1 not submitted)
S-004	B - paper fallback	0	24	Yes	Yes	N/A (paper)
S-005	D - engaged, empty	0	0	No	No	N/A (no submitted assessment)

Table 12. Full-period participation across 24 expected card-family units. All five are in N_DIARY; 4/5 provide data and 3/5 complete the full schedule. Paper contributes clinical completeness but not electronic timing.

8.2 ENDPOINT-SPECIFIC DAY-1 PAIN

For Day-1 pain, S-001 reports Grade 1 within 24 hours; S-002 reports Grade 2 at 29 hours; S-004 has a Grade-0 paper negative; and S-003 and S-005 are missing. If investigator precedence changes S-002 to Grade 3, occurrence remains positive, severity changes, and timing remains 29 hours from the linked subject entry.

USUBJID	Day-1 injection-site pain	Pain present?	In M?	In N_DIARY?	Any diary data in Days 1–8?
S-001	Grade 1 (e-Diary, ≤24 h)	Yes	Yes	Yes	Yes
S-002	Grade 2 (e-Diary, 29 h)	Yes	Yes	Yes	Yes
S-003	Missing (Days 2–8 available)	-	No	Yes	Yes
S-004	Grade 0 (paper; timing N/A)	No	Yes	Yes	Yes
S-005	Missing (no diary data)	-	No	Yes	No

Table 13. Day-1 pain. In the unrestricted primary analysis, n = 2 and M = 3. S-004 contributes a paper negative to M; S-002 enters the ≤48- and ≤72-hour views but not the ≤24-hour view.

8.3 ONE EVIDENCE PATTERN, SEVERAL QUESTIONS

The same profiles answer endpoint-incidence, participation, submission, and entry-timing questions. The numerator and denominator must be defined at the same grain.

For unrestricted Day-1 pain, 2/3 = 66.7%. Using the four subjects with any diary data during Days 1–8 gives 2/4 = 50.0%, but mixes period participation with Day-1 evaluability. Using all five expected subjects gives 2/5 = 40.0% and treats missing as no event. Separately, 4/5 = 80.0% provide any diary data and 3/5 = 60.0% complete all 24 card-family units. These are different questions, not competing estimates.

Operational rates also depend on their frames. After S-004's approved switch to paper, the current electronic-collection frame comprises S-001, S-002, S-003, and S-005; 2/4 submit Day-1 pain. Among the two submitted electronic records, 2/2 are within 72 hours and 1/2 within 24 hours. These are submission and timing metrics, not reaction incidence.

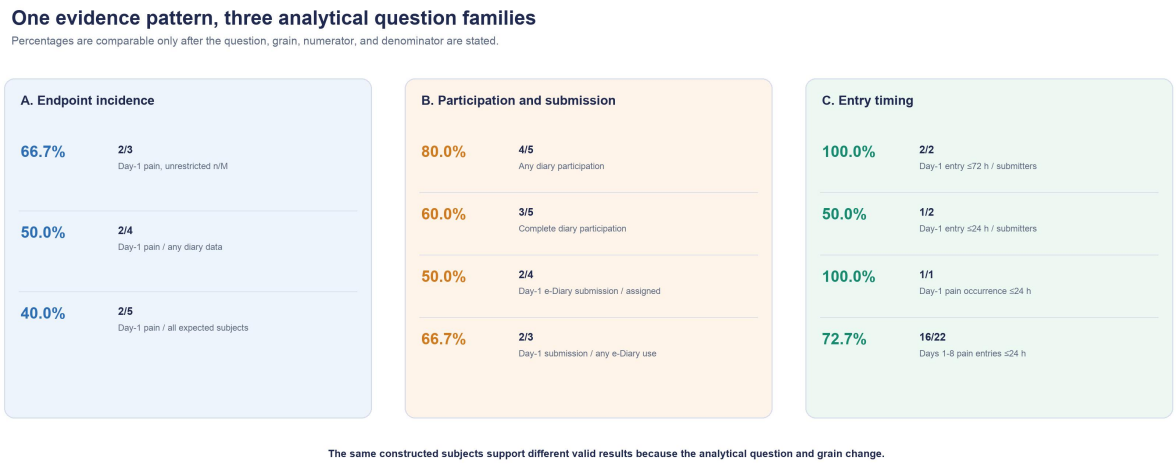

Figure 4. One evidence pattern, three question families: endpoint incidence, participation/submission, and assessment-level timing.

The ≤24-hour Day-1 pain view has n/M = 1/1 because only S-001 contributes an evaluable result in that window; this differs from 1/2 = 50.0% of submitted electronic Day-1 pain records entered within 24 hours. Across the Days 1–8 pain grid, 16/22 = 72.7% of timing-evaluable entries are within 24 hours. Every output must define its anchor, grain, numerator, and denominator.

9. FUTURE DIRECTIONS, LIMITATIONS, AND SCOPE

The same framework becomes more important as ePRO, decentralized collection, BYOD, and connected devices increase data volume and move capture toward the participant.

It can also extend beyond reactogenicity. In some vaccine efficacy trials, e-Diary symptoms trigger illness visits or testing [31]; a missed or delayed entry can then affect case ascertainment. The same principle applies in other therapeutic areas with frequent ePRO: preserve timing, missingness, source, and triggering logic so the endpoint can be reconstructed.

The proposed vocabulary, flags, diary states, and dataset labels are sponsor conventions, not CDISC-endorsed standards. The examples are constructed and do not assume that ePRO systematically raises or lowers reaction rates; the direction depends on behavior, timing, and system design.

CONCLUSION

Electronic diaries make the collection process observable. A result can carry clinical time, an intended-day anchor, trusted entry time, receipt time, collection mode, and recorder. These distinctions matter because denominators answer questions: endpoint M, participation, submission, completeness, and timing summaries use different pre-specified populations. They should be reported separately and compared across treatment groups before reaction rates are interpreted. Preserve the expected-assessment frame, source keys, timing metadata, and analysis flags needed to reproduce those choices from source through SDTM, ADaM, TFLs, and reviewer guides. Capture context plus explicit denominators turns an e-Diary card into a reproducible, submission-ready analysis result.

REFERENCES

[1] U.S. Food and Drug Administration. Clinical Outcome Assessment (COA): Frequently Asked Questions. Accessed 25 July 2026. https://www.fda.gov/about-fda/clinical-outcome-assessment-coa-frequently-asked-questions.
[2] U.S. Food and Drug Administration. Digital Health Technologies for Remote Data Acquisition in Clinical Investigations: Guidance for Industry, Investigators, and Other Stakeholders. December 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations.
[3] Clinical Data Interchange Standards Consortium (CDISC). Vaccines Therapeutic Area User Guide (TAUG-VAX), Version 1.1. 2018. https://www.cdisc.org/standards/therapeutic-areas/vaccines/.

[4] U.S. Food and Drug Administration, Center for Biologics Evaluation and Research. Submitting Study Datasets for Vaccines to the Office of Vaccines Research and Review: Guidance for Industry, Technical Specifications Document, Version 2.1. December 2019. https://www.fda.gov/media/112581/download.
[5] Huang Q, Moeller E, Morgan L, et al. A New Data Model to Meet CBER Guidance on Submission of Reactogenicity Data for Vaccine Trials. ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop; 2021.
[6] Mastri J, McLaughlin S. Adapting and Evolving with OVRR Requirements. Presented at the CDISC 2023 Europe Interchange; 2023.
[7] Clinical Data Interchange Standards Consortium (CDISC). Study Data Tabulation Model Implementation Guide: Human Clinical Trials (SDTMIG), Version 3.4, and Study Data Tabulation Model (SDTM), Version 2.0. 29 November 2021.
[8] Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials. 2003;24(2):182-199. doi:10.1016/S0197-2456(02)00320-3.
[9] Sreeraman C, George J. Journey of Reactogenicity Data from EDC to SDTM to Analysis Datasets. PHUSE US Connect 2024; Paper TA07.
[10] Delong PS, Humler D, Haag T, Yeomans A, Andrus J, Eremenco S, et al. Best Practice Recommendations for Electronic Clinical Outcome Assessment Data Changes. J Soc Clin Data Manag. 2023;2(1):12. doi:10.47912/jscdm.249.
[11] Bissonett A. A SAS®sy Study of eDiary Data. PharmaSUG 2017; Paper BB14.
[12] Skinner J, May R. Combining Electronic Diary Seizure Data with Visit-Based Seizure Data in the MONEAD Study. PharmaSUG 2018; Paper RW-02.
[13] pharmaverse. admiral: ADaM in R Asset Library [R package]. Accessed 16 July 2026. https://pharmaverse.github.io/admiral/.
[14] xportr development team. xportr: Utilities to Output CDISC SDTM/ADaM XPT Files [R package], Version 0.5.0. Accessed 16 July 2026. https://atorus-research.github.io/xportr/.
[15] International Council for Harmonisation (ICH). E9(R1) Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials. Final version. 20 November 2019.
[16] U.S. Food and Drug Administration. Submitting Patient-Reported Outcome Data in Cancer Clinical Trials: Technical Specifications Document. November 2023. https://www.fda.gov/media/173581/download.
[17] Amdal CD, Falk RS, Alanya A, et al. SISAQOL-IMI consensus-based guidelines to design, analyse, interpret, and present patient-reported outcomes in cancer clinical trials. Lancet Oncol. 2025;26(12):e683-e693. doi:10.1016/S1470-2045(25)00520-0.
[18] International Council for Harmonisation (ICH). E6(R3) Good Clinical Practice: Principles and Annex 1. Final version, 6 January 2025; error corrections, 24 October 2025. Accessed 16 July 2026. https://www.ich.org/page/efficacy-guidelines.
[19] eClinical Forum. eCF Requirements for Electronic Data for Regulated Clinical Trials. Public Release PR2023.1; 2024.
[20] European Medicines Agency. Guideline on Computerised Systems and Electronic Data in Clinical Trials. EMA/INS/GCP/112288/2023. 9 March 2023.
[21] National Medical Products Administration, Center for Drug Evaluation. Guideline for the Application of Patient-Reported Outcomes in Drug Clinical Research (Trial). 27 December 2021.
[22] National Medical Products Administration. Guideline on Real-World Evidence to Support Drug Development and Regulatory Review (Trial). Announcement No. 1 of 2020. 7 January 2020.
[23] Sugimoto N, Morimasa M, Misawa H, et al. A Preliminary Study Introducing Electronic Patient-Reported Outcome Using Bring Your Own Device in Post-marketing Surveillance in Japan. Ther Innov Regul Sci. 2026;60:199-209. doi:10.1007/s43441-025-00873-0.
[24] U.S. Food and Drug Administration. Part 11, Electronic Records; Electronic Signatures - Scope and Application: Guidance for Industry. August 2003. https://www.fda.gov/media/75414/download.
[25] U.S. Food and Drug Administration. Study Data Technical Conformance Guide. June 2026. https://www.fda.gov/media/153632/download.
[26] U.S. Food and Drug Administration. Dataset-JSON Pilot Report and Next Steps. 25 July 2024. https://www.fda.gov/drugs/news-events-human-drugs/dataset-json-pilot-report-and-next-steps-07252024.
[27] Clinical Data Interchange Standards Consortium (CDISC). Dataset-JSON, Version 1.1. Accessed 16 July 2026. https://www.cdisc.org/standards/data-exchange/dataset-json.
[28] Clinical Data Interchange Standards Consortium (CDISC). Analysis Data Model Implementation Guide (ADaMIG), Version 1.3. 29 November 2021. https://www.cdisc.org/standards/foundational/adam/adamig-v1-3.
[29] pharmaverse. admiralvaccine: Vaccine Extension for admiral [R package]. Accessed 16 July 2026. https://pharmaverse.github.io/admiralvaccine/.
[30] Minjoe S. Adding Rows to a BDS Dataset: When to Use DTYPE in Addition to Metadata. PharmaSUG 2021; Paper DS-081.
[31] ModernaTX, Inc. Statistical Analysis Plan, Version 2.0, Protocol mRNA-1273-P301. 8 March 2021. ClinicalTrials.gov Identifier NCT04470427. https://cdn.clinicaltrials.gov/large-docs/27/NCT04470427/SAP_001.pdf.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Jialiang LIU - Sanofi - Jialiang.liu@sanofi.com ; Lin CHEN - Sanofi - Lin11.chen@sanofi.com