

A closer look at data quality in the EXPAND trial

Sponsors need confidence that study data provide a coherent, dependable basis for interpretation. Achieving that requires both planned checks and the ability to investigate questions that emerge as the data accumulate. This case study explores how Nimble's AI-assisted approach can extend that review. A focused investigation of released EXPAND trial data identified 12 discrepancy types and additional concerns through clinical-context and pattern review, illustrating the value of deep AI-driven data quality review.

The study

EXPAND (NCT00678535; EMR200048-052; Lordick et al., Lancet Oncology, 2013) was a randomized, multicenter study sponsored by Merck KGaA in people with advanced gastric or gastroesophageal-junction cancer. It tested whether adding cetuximab to capecitabine-cisplatin chemotherapy improved outcomes. For this case study, we reviewed deidentified data from the chemotherapy-only control arm, made available through Project Data Sphere. Our objective was to demonstrate what a focused application of Nimble's review approach could reveal in real study data.

436 participants | 48 datasets | 237,784 rows

Review approach

We started with the study protocol, annotated CRF and data dictionary, then used them to frame practical questions about the data. AI agents autonomously explored related datasets, developed and executed checks, and investigated candidate concerns in context. Separate verification passes checked retained findings against original records; repeated issues were grouped by type and explained flags were removed.

This was a focused investigation to identify illustrative concerns, rather than a comprehensive application of our review process. We used AI-driven techniques, spending one day on planning and discovery and another on validation. The findings are organized into three evaluation categories within Nimble's review framework:

Programmed checks use defined rules and calculations to spot inconsistencies.

AI checks help make sense of information spread across related clinical records.

Pattern review compares groups of participants to identify recurring gaps, unusual distributions and differences that merit investigation in the context of the study.

Findings from programmed checks

We identified 12 types of discrepancies across five areas of the data.

Severity describes the potential impact on data review, rather than confirmed clinical harm. High indicates possible implications for safety, treatment, eligibility or response interpretation; Moderate indicates an important documentation or coverage concern; Low indicates a more limited record discrepancy.

ISSUE	COUNTS/REF.	SEVERITY	WHY IT MATTERS
Measurement integrity			
Neutrophil counts were higher than the total white-cell counts. In five matched samples, neutrophil values were hundreds of times greater than the total white-cell count, even though the units matched.	5 paired samples G130-F01	Low	These figures cannot both be correct and could distort the interpretation of safety results.
Body surface area values appeared to be on the wrong scale. Recorded values of 162-194 m ² contrasted with approximately 1.67-1.94 m ² calculated from height and weight.	4 visit records G130-F04	Low	Because body surface area informs dose calculations, the difference should be traced through treatment records.
The phosphate lower reference limit exceeded the upper limit. Seven records showed a lower reference limit above the upper limit.	7 records G130-F02	Low	A reversed range makes it hard to judge whether a laboratory result is low, normal or high.

ISSUE	COUNTS/REF.	SEVERITY	WHY IT MATTERS
Tumor & disease assessment			
Overall tumor response did not agree with the underlying assessments. Five response entries conflicted with their component results, while one response was missing despite complete components.	5 conflicts 1 missing response G130-F03	High	These inconsistencies affect how response or progression is represented in the record.
A scan was marked as not performed, although measurements were present. The same assessment included CT-based tumor measurements.	1 assessment G130-F23	Low	The procedure status and the measurement data need to be reconciled.
The total count of metastatic sites did not match the organs selected. One assessment listed fewer sites in its summary than in the detailed organ selections.	1 assessment G130-F06	Low	The summary may underrepresent the extent of disease recorded elsewhere.
More target lesions were selected than the protocol allowed. The baseline lesion set exceeded the limit specified in the supplied protocol.	1 assessment G130-F21	Moderate	This matters because the selection rule is intended to make tumor assessments consistent.
Non-target lesion response did not match the lesion history. Tracked lesions had disappeared, but the non-target response was not recorded as complete.	1 assessment G130-F22	Moderate	The response needs review, although separate evidence of new lesions supported overall progression.
Eligibility			
The eligibility decision conflicted with the screening answers. Thirteen assessments recorded overall eligibility as "Yes" even though an inclusion criterion was marked "No."	13 assessments G130-F19	High	The final eligibility status does not agree with the answers used to support it.
Treatment exposure			
The total capecitabine amount did not match the cycle plan. One cycle was marked "taken as planned," but its total was 800 mg. The protocol-based plan implied 28 administrations of 800 mg, or 22,400 mg.	1 cycle G130-F07	Moderate	The inconsistency makes the treatment exposure difficult to interpret confidently.
Record chronology			
Shipment dates came before sample collection dates. In two records, the shipment date preceded collection for the same sampling timepoint.	2 records G130-F05	Low	The recorded dates do not describe a plausible specimen timeline.
Medication end dates preceded the earliest possible start. Three records had a month-only start date and an exact end date in the preceding month. Even the first day of the recorded start month fell after the end date.	3 intervals G130-F20	Low	These dates cannot reliably establish medication duration or overlap with other events.

Findings from AI checks

Some severe blood-count episodes were not clearly reflected in the adverse-event history

Severity: Low

After taking earlier adverse events and recovery between episodes into account, four Grade 3 neutrophil episodes across three participants still warranted review.

Why it matters: these may represent separate episodes that the AE history does not fully capture. The clinical context is needed to decide whether an existing event already accounts for them.

Looking at more of the clinical history resolved most of the initial concerns

Five missed-dose flags were attributed to toxicity but initially appeared to lack supporting clinical records. Reviewing the full treatment cycle and neighboring cycles explained four of them.

Findings from pattern review

Site, country and region identifiers were unavailable in the deidentified release, so the pattern review compared groups defined by their assessment records rather than by geographic location.

ECG records showed different assessment schedules across participant groups

Severity: Moderate

Among 169 participants with at least six dated Day 8 assessment opportunities, 38 had no ECG record within the assessed Day 8 windows. Yet 37 of these participants had ECGs recorded under Day 1 visits, and 35 had them repeatedly—commonly in cycles 2, 4 and 6. Of their 111 Day 1 ECG records, 95 occurred within one day of the recorded Day 1 visit, supporting a genuine timing difference beyond visit labels alone.

In contrast, the other 131 participants had ECG records within 974 of 1,310 Day 8 windows—approximately 74%. Within this comparison group, 31 participants had both Day 1 and Day 8 ECG records; in 29, all Day 1 ECGs preceded the Day 8 ECGs, suggesting a change in assessment or recording practice during follow-up.

Why it matters: the supplied protocol requires ECGs on Day 8 of every cycle, allowing up to three days early. It does not describe a substitute schedule of Day 1 ECGs in alternate cycles. The recurring difference between groups, together with the within-participant transitions, warrants review of protocol implementation, assessment timing and recording practices. Earlier or local amendment applicability would need checking to determine the explanation.

Discussion

For this focused investigation, the review questions were autonomously developed from the study materials and the data themselves. A comprehensive review would extend coverage to additional areas and questions. This case study demonstrates value within a limited scope; it does not establish the full extent of concerns in EXPAND. A broader engagement would agree the priority areas and review depth with the sponsor, then apply the complementary methods across that scope.

The retained results span defined-rule discrepancies, contextual concerns and recurring patterns. Each calls for a different kind of evaluation. Bringing these approaches into one review gives sponsors a broader view of data quality and a common basis for prioritizing concerns across clinical, data management and biometrics teams.

The investigation produced traceable evidence and explanations alongside the concerns, and removed initial concerns that further investigation resolved. That makes the output a basis for further assessment. In a Nimble engagement, the findings report, check register and supporting evidence give the sponsor a defined body of work to review and a basis for deciding where follow-up is warranted.

Some findings show direct inconsistencies; others point to questions that need further review. Their severity ratings are subjective and indicate potential importance for the study objectives. Masked dates, suppressed narratives and the absence of audit history and final document versions limit what can be concluded.

The 12 discrepancy types account for 45 logical observations, although these are measured in different units (for example, sample pairs, assessments and date intervals). Additional signals may overlap with existing types and are not included in that figure. We are not presenting a detection rate or a deduplicated count of all checks.

For sponsors, that capability offers a way to extend review as a study evolves: emerging questions can become additional checks and investigations. Review depth can grow with the available evidence and study priorities.

A focused application of Nimble's AI-assisted approach surfaced a varied set of evidence-backed concerns in released data from a completed trial. For sponsors, the opportunity is to extend the depth of study review and direct expert attention toward questions supported by the data. This case study provides a concrete example of that value.

Sources and acknowledgments

Project Data Sphere, EXPAND package 130 | NCT00678535 / EMR200048-052 | Lordick et al., Lancet Oncology, 2013

Study materials: protocol v5.0, annotated CRF and data dictionary.

This publication is based on research using information obtained from data.projectdatasphere.org which is maintained by Project Data Sphere. Neither Project Data Sphere nor the owner(s) of any information from the web site have contributed to, approved or are in any way responsible for the contents of this publication.