Automating Clinical Trial Data Collection to Improve Data Quality: The SWOG–nCartes Collaboration

Chris Cook, Keith Goodman, Sunita Yadav, John McIlwain, Antje Hoering, and Michael LeBlanc

SWOG CANCER RESEARCH NETWORK

Abstract

Collecting clinical trial data often requires duplicative and time-consuming manual entry. Clinical information must first be entered into the medical record and then re-entered into a separate electronic data capture (EDC) system structured specifically for research activities. This process is labor-intensive, costly, and prone to transcription errors. Widespread adoption of automated approaches has been limited by challenges in harmonizing data structures across diverse clinical sites and the lack of universally implemented standards.

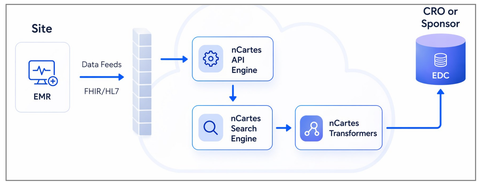
To address these barriers, the SWOG Cancer Research Network has partnered with nCoup over the past four years in utilizing nCartes, a cloud-based software platform designed to automate the transfer of clinical data from site electronic medical record systems to research databases. Automating this process reduces staff workload and reduces common sources of manual transcription errors, thereby improving data quality and lowering downstream data-cleaning costs.

Since May 18, 2022, SWOG has evaluated the impact of nCartes on data quality and operational efficiency in the S1803 multicenter myeloma trial (DRAMMATIC study: Phase III Study of Daratumumab (NSC-791647) + Lenalidomide (LD) or Lenalidomide (L) as Post-Autologous Stem Cell Transplant Maintenance Therapy in Patients with Multiple Myeloma Using Minimal Residual Disease to Direct Therapy Duration). In an institution-matched analysis comparing pre- and post-nCartes implementation, we reviewed over 3,300 data queries across 7 sites and 81 participants. As of November 2025, sites that consistently implement nCartes have seen substantial reductions in query rates, including near-complete elimination of structured-data queries, reflecting significant improvements in data accuracy and completeness. These gains translated to decreased effort for data managers and site staff and increased likelihood of delivering a high-quality dataset at trial closeout.

Our findings support automated data transfer as a promising and scalable strategy for improving data quality, reducing site burden, and strengthening the overall efficiency of multicenter clinical trials.

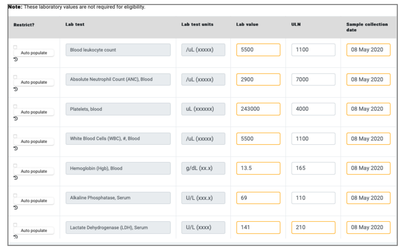
Background

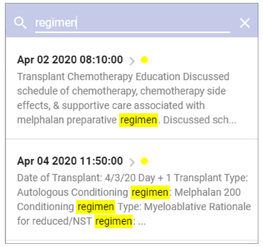
An ideal data entry model is that data are entered once and that data can be easily reused downstream in other systems without any duplicative manual data entry. The SWOG Cancer Research Network has been investigating ways to enhance and automate data entry to help improve site efficiency and overall data quality. In 2021, SWOG partnered with nCoup to deploy their nCartes system at SWOG sites to help advance data collection. With nCartes, the site enters their data once into the Electronic Medical Record (EMR) and the data (e.g., labs, medications) are automatically fed to the nCartes system and prepared for the EDC. The systems diagram below shows how data feeds into nCartes which processes most of the data transfer instead of a study coordinator manually copying data to the EDC system.



Data that are already well structured in the EMR system have the most potential to see efficiency gains. These data include demographics, vital signs, labs, and medications.

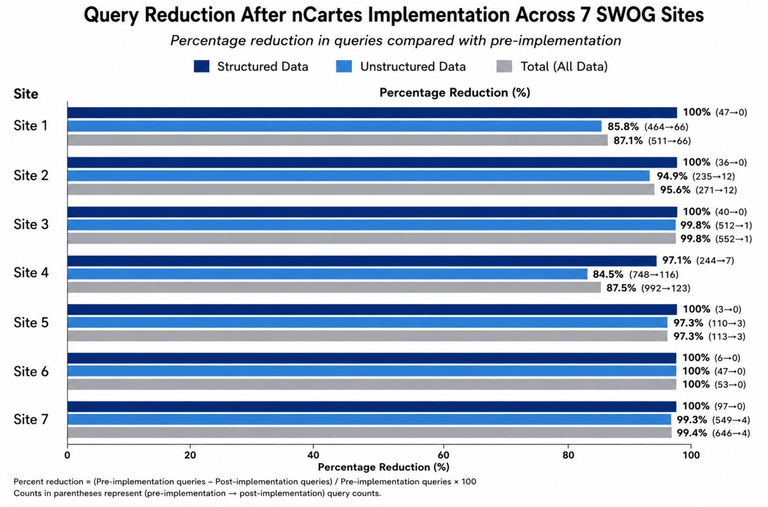
While the nCartes system is optimized for structured data similar to the lab form example below on the left, the nCartes system also allows sites to fill in any additional unstructured data. Reporting of some unstructured data (e.g., progress notes) is simplified through the nCartes' search tool which is depicted below on the right. Within nCartes, users can search the progress notes to look up information more easily and/or copy and paste information from the notes.





Data Quality Impact

To evaluate changes in data quality before and after nCartes implementation, we measured data query rates, which are typically generated when data quality issues are identified. In an institution-matched analysis comparing pre- and post-nCartes implementation, we reviewed more than 3,300 data queries across 7 sites and 81 participants. As of November 2025, all 7 sites demonstrated reductions in query rates following implementation of nCartes. Query reductions ranged from 97.1% to 100% for structured data and from 84.5% to 100% for unstructured data. The results are summarized in the figure below.



Query reduction rates were higher for structured data, consistent with expectations for the nCartes system. Structured data are stored in predefined formats, making them easily searchable, machine-readable, and well suited for automated transfer with minimal error. Examples include demographic data (e.g., age) and vital signs (e.g., blood pressure). Systems such as nCartes are specifically designed to automate transfer of these data and can nearly eliminate manual entry errors.

Substantial reductions were also observed for unstructured data. Unlike structured data, unstructured data do not follow predefined formats and often require manual interpretation or review. Examples include clinician notes and radiology reports. In this study, adverse event data were also considered unstructured because the number and type of adverse events could not be predefined in advance for structured transfer.

Next Steps

The results of this analysis indicate that the nCartes system substantially reduces data entry errors across both structured and unstructured data. These reductions improve overall data quality while decreasing staff time spent resolving queries. For example, if similar ~90% query reduction rates were achieved across the full study, more than 1,000 hours of query resolution time could be saved, using conservative assumptions that sites require an average of only 2 minutes to resolve each query.

One limitation of this study is its pre-post design, as other factors may have contributed to reductions in query rates over time. However, given both the magnitude of the reductions observed and their temporal association with nCartes implementation, nCartes was likely the primary driver of the observed improvements in data quality. We plan to expand both implementation and analysis efforts to further validate these findings across additional studies and datasets.

Another limitation is that only a small proportion of the submitted data were structured; approximately 12% of submitted data fell into this category. Increasing the proportion of structured data is an active area of development. One such effort involves systematic conversion and transfer of Adverse Event (AE) data into structured formats.

Additional development efforts focus on improving automation for unstructured data processing. For example, ECOG Performance Status may appear in clinical documentation under several synonymous terms, including “ECOG Score,” “Performance Status,” and “Zubrod Score.” Semantic standardization enables these terms to be automatically recognized and mapped to a single standardized concept. This approach improves transparency, reduces ambiguity, and promotes consistent data capture when clinical documentation differs from protocol terminology.

We plan to report on these ongoing efforts in future publications.

Acknowledgements

This work was supported by SWOG NCI/NCTN grants U10CA180888 and U10CA180819. We thank the participating sites and SWOG for their support.

nCartes

CRAB Cancer Research And Biostatistics

Fred Hutch Cancer Center