Paper DH10

Enhancing Quality of External Data

Pascale Youinou, Business & Decision Life Sciences, Brussels, Belgium
Kimberly Devis, Business & Decision Life Sciences, Brussels, Belgium

ABSTRACT

During clinical study set-up, both sponsor and CRO have a lot to consider, especially when collaborating with many different vendors (EDC, IWRS, ePRO, etc.). Sponsors still may choose not to integrate all databases in one system. When your normal process considers having DM and Stat teams involved upfront to proactively minimize the issues for data cleaning, SDTM, analysis and submission, working with primary output external to EDC might affect data quality. Often multiple vendors participate in the study and the DM team must deal with several databases' reconciliation in parallel. In this presentation we will share our experience with:

- Parallel set-up of several databases
- Discrepancies in data collection and protocol understanding
- Peculiarities of different external reconciliations
- Integration Pros and Cons
- Quality challenges for late reconciliation (deviations, knowledge of each DB specificity, SDTM/ADaM impacts, etc.)
- Communication challenges, management of teams from different time zones.

INTRODUCTION

Significant amounts of data are collected for supporting clinical study design endpoints originating from a variety of sources. Clinical data collection often involves several databases, including the electronic data capture (EDC) system, Interactive Response Technology (IRT) systems, electronic Clinical Outcome Assessment (eCOA), and other external vendor databases. The EDC system typically comprises electronic case report forms (eCRFs) created by the Clinical Data Management (CDM) team. External vendor databases are maintained by vendors who support study-specific data needs, including primary endpoints. These vendors conduct specialized testing and assessments, such as but not limited to laboratory tests, electrocardiograms (ECGs), electronic Clinical Outcome Assessment (eCOA), magnetic resonance imaging (MRI), etc.

To optimize the efficiency and data quality of the clinical database, sponsors must make crucial decisions. They can choose to integrate all databases into a unified system or maintain them separately. Additionally, sponsors should consider involving Data Management (DM), Data Standard (DS), and Statistical Programming (Stat) teams from the beginning to proactively address possible issues related to data cleaning, SDTM (Study Data Tabulation Model) preparation, statistical analysis, and submission. Alternatively, they may choose not to involve these teams initially. Furthermore, sponsors can decide to have dedicated internal teams or have them managed by one or different CROs (Contract Research Organizations).

Assuming that the DM study is overseen by a Contract Research Organization (CRO), sponsors must also decide whether the Data Transfer Specifications (DTS) process will be led by an external vendor or by the CRO CDM team. These strategic decisions significantly influence the overall success of clinical data integration and reconciliation.

In this paper, we will share our experience, as CRO, with two similar biometrics projects that involved EDC and external vendor databases. These projects illustrate the different sponsors' approaches for integration and reconciliation as well as the level of involvement of sponsors' team in the different steps (e.g., review, testing,...). We will evaluate the pros and cons of the different approaches.

STUDY SET-UP

During the development of a clinical trial protocol, study endpoints are meticulously defined, guiding the necessary data collection for the study. Typically, a CRO is responsible for building the clinical database (CDB), which captures essential information such as clinical evaluations, adverse events, concomitant medications, etc.

However, additional vendors may be contracted to meet specific study data requirements. While these vendors provide specialized services, their involvement introduces challenges related to data integration. A critical question

arises: How will data from external sources—beyond the clinical database—be collected, stored, transferred, and integrated?

Sponsors face strategic decisions regarding database integration:

- Integrated Databases: Some sponsors opt to integrate all databases into the EDC system. This approach streamlines data flow and ensures consistency. For instance, the coding tool is easily integrated. Also, IWRS (Interactive Web Response System) used to be integrated through API (Application Programming Interface). Data are synchronized automatically every few minutes and are directly accessible into EDC.

- Regular Uploads: Other databases are set to automatically upload into the EDC system regularly. With upload through sFTP server into EDC system (ex: eCOA), data are transmitted electronically on a regular basis, and EDC data are refreshed with external data immediately after (ex: daily). This simplifies data management.

- Late Reconciliation: Certain databases are transferred outside the EDC system, reconciled at a later stage and then merged with the clinical database. This approach allows flexibility but requires careful coordination. Typical examples can be the Pharmacokinetics (PK) Data.

Each strategy has its advantages and disadvantages, impacting data quality, efficiency, and overall study success.

STUDY 1, INTEGRATION, UPLOAD AND RECONCILIATION.

In the first project, a vendor database (such as IRT and coding tool) was successfully integrated into the EDC system, while some other databases were planned for regular uploads (eCOA) and reconciliations with EDC data (such as ECG, PK, etc.). To ensure a seamless parallel setup, the sponsor made a strategic decision to involve all relevant parties during the study's initiation and alignment stages.

Timelines for User Requirement Specifications (URS) and Data Transfer Specifications (DTS) creation and review were carefully coordinated among the sponsor, CRO and vendors. The EDC go-live date was synchronized with the integration of IRT and eCOA systems, allowing corresponding User Acceptance Tests (UATs) to take place in parallel. Understanding each vendor's peculiarities was essential for this accurate data integration.

Before creating any databases (both EDC and vendor databases), the team also meticulously defined key variable values. Given the trial design, a unique identification was required for subject identifiers, age/birth date, and visit identifiers. Ensuring consistent timepoints' nomenclature between the clinical database and each applicable vendor dataset resulted in efficiencies during programming and further data integration and reconciliation. For instance, because the visit names matched between the clinical and vendor databases, merging datasets required minimal effort.

Another example, maintaining a consistent subject identifier format across all supporting databases was critical. While this may seem straightforward, harmonizing identifiers becomes essential in a clinical trial involving multiple vendors and not so easy to implement. Overlooking such details during study set-up can cause complications at a later stage. Therefore, establishing a identifier format that was unique and compatible across all databases was advisable.

The advantages of aligning all partners in an early set-up stage can be summarized as follows:

1. **Efficient Data Integration**: Meticulously defining key variable values and ensuring consistent timepoints' nomenclature between the clinical database and vendor datasets streamlined the integration process.

2. **Strategic Alignment**: Coordinating timelines for URS and DTS reviews ensured a smooth integration during system's go-live phase and gain of time to avoid going back and forth.

3. **Consistency on critical variables**: Maintaining a consistent format across all supporting databases was critical. Overlooking this detail could lead to complications later.

4. **Effective Communication**: Managing parallel setups with multiple vendors requires clear communication, starting with Kick Off Meeting (KOM) between all parties (sponsors, vendors and CRO). Regular meetings and strategies for handling global time zones were crucial for successful collaboration.

5. **Attention to Detail**: Understanding each vendor's processes and aligning them with project timelines contributed to successful data integration.

6. **Quality**: Anticipating potential data collection issues ensured global data quality throughout the study.

The main challenges primarily centered around communication and technical limitations:

1. **Multiple Vendors:** Managing parallel database setups with multiple vendors having their own process and way of working required effective communication. Regular meetings involving all parties or subsets of them facilitated coordination, progress, and goals alignment.

2. **Integration environment properly configured**: Misconfiguration can lead to data transmission issues, delays impacting overall project timelines. Ensuring the environment is correctly configured before implementing integrations through API or sFTP for regular uploads is therefore crucial..

3. **Strict Timelines:** Understanding each vendor's processes and technical restrictions while aligning them with project timelines was challenging. Clear communication and flexibility were essential.

4. **Global Time zones:** Working across different time zones posed logistical challenges. Strategies for managing teams in diverse time zones were crucial for successful collaboration. For instance, practicing time zone Etiquette, such as respecting non-working hours and avoiding scheduling meetings during inconvenient timeslots was important.

Overall, initiative-taking planning, clear communication, flexibility and attention to detail were key to overcoming these challenges and achieving successful data integration.

STUDY 2, NO INTEGRATION, ONLY RECONCILIATIONS.

In the second project, no vendors were integrated into the EDC system. Instead, all external vendor data, such as IRT, ECG and PK were scheduled to be reconciled with the EDC data after the GO LIVE phase. Consequently, the set-up of all systems, including the EDC, IRT and other vendor systems, was conducted independently.

In this scenario, the sponsor does not need to ensure the close involvement of all relevant parties during the study's initiation stage. Sponsor Subject Matter Experts (SME) were assigned to each external team, as such a sponsor medical imaging expert was responsible for an MRI database set-up; a sponsor lead data manager was dedicated to the EDC overview, etc. This separation by domain expertise ensured an easy communication between SME and vendor team.

Additionally, the creation and review of timelines for URS do not require vigilant coordination among all parties involved to synchronize the GO LIVE process. Each system, while adhering to the overall project timelines, should focus on its specific development requirements.

Once all systems have been deployed into production, each reconciliation process was discussed and agreed separately between DM CRO, vendor team and sponsor DM expert. Following this agreement, reconciliation activities can be conducted extensively using the most recent and accurate data extractions throughout the study.

Late reconciliation in clinical trials has clear benefits:

1. **Efficient Set-up**: Rapid deployment of systems (such as EDC, IRT, etc.) during the study set-up prevents interaction issues, enabling focus on patient enrollment without any delay.

2. **Flexibility**: Late reconciliation allows adjustments on data collection, accommodating any unforeseen issues or discrepancies after system’s go-live. It provides flexibility in correcting errors or addressing deviations.

3. **Expertise**: SME knowledge in specific medical area or process ensures good understanding of potential issues.

4. **Picture at Cut-off Dates**: Late reconciliation captures real-world data, reflecting the actual study conduct at specific cut-off dates. This can enhance the accuracy and relevance of the datasets for interim analysis.

However, several challenges emerged during the data reconciliation phase:

1. **Vendor URS – EDC specifications incompatibilities:** The vendor database documentation was not shared with the CRO responsible for the EDC database. The vendor URS were not carefully reviewed upfront by sponsor SME, paying attention to potential format or field inconsistencies with EDC. A clear example was missing visits.

Expected visits were not included in the vendor databases, despite being essential according to EDC requirements.

Visit Name (EDC)	Visit Name (ECG)	Visit Name (Lab)
VISIT 1- SCREENING		
VISIT 2 - VISIT 1	VISIT 2	FIRST VISIT
VISIT 3 - SAFETY FOLLOW-UP	VISIT 3	SAFETY FOLLOW-UP
VISIT 4 – VISIT 2		SECOND VISIT
VISIT 5 – SAFETY FOLLOW-UP		SECOND SAFETY FOLLOW-UP
	UNSCHEDULED VISIT	UNSCHEDULED VISIT

This situation necessitated updates to the documentation and discussions during the initial DTS Creation. The DTS creation process was significantly prolonged due to extensive discussions with the sponsor and vendors to address discrepancies in database variables that were not aligned.

- **Discrepancies in Data Collection and Protocol Understanding:** Variations in data collection methods and interpretation of the study protocol led to inconsistencies. One significant issue was the **misinterpretation of key data**, such as the recording of age or date of birth, which resulted in mismatches and cases that could not be reconciled.

EDC	VENDOR
Adult patient: Age (in years): |___|___|___| Pediatric patient: Age: |___|___| |___|___| Years and Months For pediatric patient under 2 years old (from birth to 23 months of age): Date of Birth: |___|___| / |___|___|___| / |___|___|___|___| Day Month Year	Date of birth For subjects with Age>17, only year of birth is required. Enter 01 for day (DD) and JAN for the month (MMM). (eg:01-JAN-1989) For subjects with Age≥2 and ≤17, both month (MMM) and year (YYYY) are required and enter 01 for day (DD) (eg:01-AUG-2014) For subjects with Age <2, full date of birth is required (eg: 17-JUN-2022)

As a result; specific rules were established for reconciliation:

- *For patients >= 17 years and < 18 years: Do not consider month of birth during reconciliation. Check the flagged cases and put status to 'CONFIRMED' once checked and ok, else submit query.*
- *For patients>= 2 years and < 17 years: Post query into EDC when difference >= 1 month asking site investigator to clarify EDC age and vendor Date of Birth.*
- *For patients <2 years, exact match (complete Date of Birth) expected.*

- **Unclear Formats**: The date formats used by the vendor (e.g., birth dates, procedure dates) were not clearly specified. Typical examples are formats like YYYY-MM-DD, YYYY-DD-MM or DD-MM-YYYY that can easily be mixed if standard ISO format is not used. As a result, additional programming and testing steps were required to ensure a proper and consistent data reconciliation.

- **IRT Subject Mismatch**: Issues could arise with subject matching in the IRT System. Subjects are typically reconciled based on three key data points: Subject ID (SUBJID), Date of Birth, and Gender. If the format of any of these data points is inconsistent or incorrect, it can significantly impact both the programming logic and the reconciliation output

2. **Lack of Communication between sponsor and their vendors:** The sponsor anticipated that the vendor database would align perfectly with the EDC, despite not having any documentation provided or proper understanding of database specificities that can affect the other systems. Each vendor's database had unique features, requiring careful alignment with EDC requirements. Using the specific vendor DTS would have helped to understand their database structure and process. Planning a meeting with vendors to understand how they manage discrepancies between systems and their correction could help mutual understanding and speed up the problems' resolution.

   Although the initial DTS was designed to be SDTM-compatible, the sponsor requirement for the exact match with the vendor database led to further updates after the DTS was approved and further misalignment with SDTM and ADaM (Analysis Data Model) standards requirements.

3. **Multiple DTS Updates:** Frequent revisions to the DTS were required after test transfers. Dummy data used for test transfers did not include all scenarios, resulting in undetected mismatches that only became apparent once the reconciliation process began with actual production data. Consequently, this led to the need for repeated updates to the DTS to address these unforeseen discrepancies.

4. **Deviations:** Timing of reconciliation affected protocol deviations detection, not allowing prompt corrective actions and impacting interim analysis and dry-runs. Handling deviations became more complex. Defining the best way to correct the data and whose responsibility it was (the site or the vendor) was challenging.

5. **Discrepancies management by external vendors** is also something to be mentioned in the reconciliation challenges. The responsibility definition for updating the applicable databases (investigator or vendor) affected the time for resolution and issues were not fixed prior to next reconciliation run, resulting in poor time efficiency. This could become critical approaching clinical database lock. Having or not having access to vendor websites is key to fastening the cleaning. Also, the accuracy and availability of information on vendor websites is of utmost importance. The alignment of vendor database with their data transfer should be done in real-time.

In summary, effective communication, clear expectations and shared documentation are critical in overcoming these challenges during data reconciliation.

CONCLUSION

Sponsors approach to managing external databases impacts data quality. Successful data integration requires initiative-taking planning, clear communication, and meticulous attention to details during study set-up. Enhancing data quality in cases of late reconciliation requires additional cleaning and programming efforts throughout the study. In all scenarios, effective and transparent communication among all parties remains crucial.

It is important to realize that each stakeholder (CROs/vendors/sponsors) has its own process. To foster effective collaboration, it is essential to establish a mutual understanding by engaging in thorough discussions during the study set-up or Kick-Off Meeting (KOM) before initiating any activities. Understanding each party's way of working and identifying the potential differences/gaps early on can help anticipating issues and allow for adjustments in the planning (e.g., including people in reviews, being clear on test transfer expectation both with regards to timelines and to content). Persuasive communication and alignment among the parties are keys to a successful partnership.

These considerations should play a significant role in vendor selection and in defining the external data management approach, whether it involves integration or reconciliation.

CONTACT INFORMATION

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

Author Name: Pascale Youinou
Company: Business and Decision Life Sciences
Address: Av. du Bourget 3, 1140 Evere
City / Postcode: 1140 Evere, Brussels
Work Phone: +32 485 455 802
Email: pascale.youinou@businessdecision.be

Author Name: Kimberly Devis
Company: Business and Decision Life Sciences
Address: Av. du Bourget 3, 1140 Evere
City / Postcode: 1140 Evere, Brussels

Work Phone: +32 477 760699
Email: kimberly.devis@businessdecision.be