Received: 11 May 2022 | Revised: 13 August 2022 | Accepted: 19 October 2022

DOI: 10.1002/hpm.3592

RESEARCH ARTICLE

WILEY

Understanding paediatric data standards challenges through academia-industry partnerships: A conect4children (c4c) qualitative study

Anando Sen[1] | Avril Palmeri[1] | Joanne Lee[1] | Victoria Hedley[1] | Jacques Thuet[2] | Perrine Lignon[2] | Valerie Cotonnec[2] | Rebecca Leary[1] | Sinéad Nally[3] | Volker Straub[1]

[1]John Walton Muscular Dystrophy Research Centre, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

[2]Institut de Recherches Internationales Servier, Suresnes, France

[3]Novartis Pharmaceuticals, Dublin, Ireland

Correspondence
Rebecca Leary, John Walton Muscular Dystrophy Research Centre, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK.
Email: Becca.Leary@newcastle.ac.uk

Funding information
Innovative Medicines Initiative, Grant/Award Number: 777389

Abstract

Introduction: The conect4children (c4c) consortium was setup to facilitate the development of new drugs and therapies for paediatric populations and address key challenges associated with paediatric clinical trials. Two of the major adopting principles for c4c were academia-industry partnership and data harmonisation and interoperability through common eCRF definitions. To understand the challenges arising out of these principles, the c4c team at Newcastle University conducted semi-structured interviews with four c4c industry partners.

Methods: Each partner was asked 10 questions about the data standards used in their company, management and maintenance of data dictionaries, how they dealt with paediatric-specific issues, major knowledge gaps and how academia could aid in bridging these gaps. Thematic analysis was performed to identify patterns in their answers.

Results: All companies use the Clinical Data Interchange Standards Consortium (CDISC) standards but face problems when certain terminology is not included in CDISC

(e.g., paediatric-specific terminologies). All companies were committed to interoperability and had strict policies about how additional terminology could be added to their dictionaries. Three of the four companies maintained a single dictionary but also had lighter versions for specific usage. The two major knowledge gaps identified from the interviews were handling of non-CDISC terminology and maintenance of normal lab ranges in dictionaries.

Discussion: To address these gaps, c4c has been working on a four-point plan including the development of a cross-cutting paediatric dictionary and a paediatric user guide in collaboration with CDISC.

KEYWORDS
data standards, interoperability, paediatric clinical trials

Key points

- Academia-industry collaborations are crucial for paediatric clinical trials.
- Pharmaceutical companies committed to interoperability.
- Non-CDISC data items challenge interoperability between and within companies.
- c4c is developing resources to aid interoperability through collaborations.

1 | INTRODUCTION

Paediatric diseases have received increased attention in recent years due to greater public and regulatory authority awareness; however, the conduct of paediatric clinical trials faces significant challenges. The 10-year report of the Paediatric Regulation (a European Union regulation that aims to facilitate the development and availability of medicines for children, hence improving their health) identified regulatory timelines, specific populations, age-appropriate formulations and dosages of drugs, and endpoint definitions as major challenges.[1] A retrospective study analysed 326 paediatric trials under the paediatric regulation in the 2010–2014 period and found half of them required extensions while over 60% required modifications to the paediatric investigation plan (PIP).[2] These delays are attributed to setting up site-specific contracts, different ages of consent in different countries and obtaining parental consent. Several studies have discussed recruitment challenges in paediatric clinical trials.[3,4] There has been a general reluctance both on the part of parents and doctors to enrol children in trials because of fears of possible harm. Metabolic and biological changes in children makes each cohort (e.g., neonates, infants, etc.) very specific and difficult to account for their growth as they age throughout the trial. Drug formulations for younger children require palatable forms since oral routes (including ability to swallow and tablet size) may not be possible. Intravenous or intramuscular routes have ethical issues as they may induce fear (of pain). Dosages need to be continuously updated and are dependent on the physical development of the child. Such development also hinders the definition of an outcome or endpoint. For example, the perception of pain or nausea can be different at the start and end of a trial. While some alternatives (e.g., face pain scale[5]) have been suggested, these are subjective outcomes and raise other issues.[6,7] Many of the

challenges inherent to planning and conducting paediatric research are shared by the rare disease community; this is unsurprising, as most paediatric conditions are in fact classified as rare (i.e., a disease that affects less than 200,000 individuals in the USA or less than one in 2000 people in Europe).[8,9] For both communities, multistakeholder collaborations are key to more focussed, integrated research. There are also strong similarities in terms of the data needs of both populations: with small, geographically separate populations, data is particularly precious.

To address these challenges a consortium, Collaborative Network for European Clinical Trials for Children (conect4children, or c4c) was set up under the Innovative Medicines Initiative (IMI) 2 Joint Undertaking. With a streamlined approach c4c aims to provide a consistent framework for drug development, avoid logistical delays by using the same systems and promote collaboration between scientific entities (particularly academia and industry).[10–12] Overall, the c4c network includes 35 academic institutions, 10 industry partners from the European Federation of Pharmaceutical Industries and Associations (pharmaceutical companies), 50 third party partners and around 500 affiliated partners. These are based in 21 countries and each country has a National Hub (Finland and Iceland share a national hub).[13] Each National Hub is linked to several hospital sites and coordinates activities for high-quality clinical trials to take place. The National Hubs along with the c4c network as a whole support these sites in the design and execution of clinical trials, with standardized procedures and training. It must be noted that c4c cannot address every aspect of paediatric clinical trials. For example, it has no control of paediatric regulations, institutional review boards or ethical standards of different entities.

A review article describing the c4c initiative, laid out nine adopting principles.[14] While detailed discussions on each of these nine principles are beyond the scope of this paper, we focus on two of the principles—academia-industry partnership and data interoperability. Collaboration between academia and industry has been viewed positively in several healthcare applications.[15–17] Generally, academia (including the linked sites) provides access to patient populations, expertise in the clinical management of paediatric disease and, healthcare facilities, while industry provides expertise in the drug development pathway and clinical trial sponsorship. However, since the objectives of academic and industry institutions are different, there may be differences of opinions—particularly in publications. Historically there has been a reluctance on publishing negative results though it is now universally agreed that negative results have scientific value, and it is a good scientific practice to publicize them.[18,19] Prominent authorships positions - first and corresponding, can be another source of differences. Hence, there is an urgent need for academic and industry partners to understand each other's concerns through direct engagement. There are also opportunities for industry and academics to learn from each other's experiences and solve the more complex issues relating to paediatric clinical research together.

Data interoperability is critical for bringing data into a common format. For paediatric trials, c4c aims achieve interoperability through standardized case report forms (CRFs) across institutions. This in turn will facilitate collaborative research across institutions by bringing together the data from different sources. Pooling together multi-institutional data is particularly important for rare disease, where each individual institution's data may lack statistical power for any meaningful analysis. It is imperative that data of benefit for research in rare and paediatric conditions is collected and standardized in a manner which allows for secondary use, wherever possible, to support further research in the future. An important tool to standardize and harmonize data is a data dictionary. While numerous data dictionaries have been developed at institutional levels, there has been a push to develop common data standards at higher levels. Two such standards are the Common Data Model (CDM)[20] by Observational Health Data Sciences and Informatics (OHDSI) and the Operational Data Model (ODM)[21] by Clinical Data Interchange Standards Consortium (CDISC). The CDM was originally developed by the Observational Medical Outcomes Partnership (OMOP) and carried forward by OHDSI. The CDM facilitates analysis of different databases by transforming data into a common format and representation, and then performing analyses using routines based on the common format. The ODM, which is independent of vendors and platforms, is used for the exchange and storage of clinical and translational research data, along with the associated metadata. The CDISC ODM has become the preferred standard for representing CRFs content in many electronic data capture tools. This is largely due to regulators—Food and Drug Administration (FDA) in the United States and Pharmaceuticals and Medical Devices Agency (PMDA) in Japan, mandating the use of CDISC

standards for regulatory submissions.[22–24] Despite such large data models being available, concerns remain whether they capture the paediatric terminologies sufficiently.[25,26]

To address these issues the c4c team at Newcastle University conducted a series of interviews with representatives of c4c industry partners. The major goals of these interviews were to understand the data standards used by these companies, how they are managed and maintained, their experience with paediatric-specific data standards, the knowledge gaps they are facing, and the ways academia could collaborate with them to bridge these gaps. This paper reports the findings from these interviews and the ongoing efforts to address the concerns expressed.

2 | METHODS

2.1 | Organisation

The research within c4c is overseen by a project leadership team and is organised into eight work packages. There is a work package (WP5) dedicated to standardisation and harmonisation of paediatric data. Newcastle University (NU) is a co-lead of WP5 and decided to conduct interviews with industry partners (also a part of WP5). An 'industry research working group' (IRWG) was formed to take this forward (RL, JL, AP, VH). The IRWG decided the questions for the interviews based on the adopting principles of c4c (particularly interoperability and academia-industry partnerships). Four industry partners were invited for the interviews and all accepted.

2.2 | The interviews

Due to confidentiality agreements the names of the companies cannot be disclosed. We will refer to them as Company A, B, C and D. The companies are large multinational companies with tens of thousands of worldwide employees. Since c4c is a European project, the companies' European locations are part of c4c. Each company was provided with the list of questions in advance of the call to help them prepare, and to allow them to identify the most appropriate personnel (e.g., members of the data management/standards development teams) to join the call. While the job titles of the interviewees cannot be disclosed, they were all in leadership or senior positions.

The interviews were semi-structured and conducted online (over Microsoft Teams) by the Newcastle University team (RL, JL and AP). Each interview started with a presentation (by RL) about the c4c data tasks, to provide context. One member of the Newcastle University team (either AP or JL) then asked the questions. If further information or clarifications were required, additional questions could be asked by all members of the team. Each call was recorded, and notes were taken (either by AP or JL). These notes were formalised into reports after the interviews. During the interview, if the company representative(s) was unable to provide an answer, they were allowed to get back to the NU team later via email. While no explicit instruction was provided for answering the question in a global context, given the pan-European nature of c4c, it would be safe to assume the answers would at least apply to all of Europe. All company representatives were notified that in case of any publication arising from the interview, they would have the opportunity to provide or decline consent, and to review the manuscript.

The following questions (in the same order) were asked to the company representatives.

1. What is the format of your Data Dictionary/Dictionaries? *(PDF/Excel/online tool/other)*
2. How are these dictionaries used?
3. Is there one data dictionary per company, or does it vary from study to study?
4. Who maintains the data dictionary? *(Job titles/full-time or part-time employees/what are their skillsets?)*
5. How is consensus reached on how to represent data items? *(Internal standards/using CDISC methodology/other)*
6. Does the format result in interoperable data between studies?

7. How long does it typically take to develop a data dictionary? *(For example, if creating a dictionary for a new therapeutic area)*
8. What is the cost to develop and maintain a data dictionary? *(Maintenance per year)*
9. Are there any gaps in knowledge relating to data harmonisation that C4C could potentially fill? *(Through training, expert advice or consultancy)*
10. Which paediatric data items represent "pain points"? *(Please tell us about any paediatric data items that cause you significant difficulty, we will consider them for the tools we are developing)*

2.3 | Data analysis

Question-wise thematic analysis was performed on the interview reports (by AS) using the methodology described by Braun and Clarke.[27] Questions 9 and 10 were combined for the analysis as they dealt with similar themes. The reports were read several times for familiarisation. Following this, the initial annotations (or codes) were generated. An annotation is a basic unit from the answer text that can be meaningfully analysed and is relevant to the corresponding question. They could be words, phrases or even sentences. There was no limit on the number of annotations per answer. For follow-up questions, the same annotation appearing in both the question and answer was annotated once. These annotations were then grouped into themes based on their conceptual similarity. The themes were reviewed to ensure all annotations within a theme were appropriate, no new themes could emerge from the answers and patterns were then identified. Themes of particular interest were the ones that contained annotations from the answers of multiple companies. Lastly, the themes were named ensuring precise representation of the annotations from which they were generated. For validation, the themes and annotated reports were presented to an independent author (AP), who had no role in the thematic analysis up to this point and critically analysed until consensus was reached.

3 | RESULTS

Companies A, B C and D had 1, 2, 3 and 2 representatives that were part of the interview and provided answers. Company B followed up with email responses for questions 9 and 10. The important points about their responses are tabulated in Table 1.

3.1 | Thematic analysis

The themes that contained annotations from more than one company are listed in Table 2 along with the total number of annotations and example annotations. An example of the annotation process is shown in Figure 1 for question 1 asked to company B. While themes that contained annotations from just one company are not a part of Table 2, they can be found in Table 1 and in the discussion below.

All four companies maintain data dictionaries in Excel though two of the companies (A and B) are transitioning to a metadata repository (MDR). The repository used by company A will be searchable and could allow for a high-level of automation by adding standards to their Electronic Data Capture (EDC) system. The MDR tool for company B will have the capability to export to Excel, if needed. When discussing development of advanced tools, Company D suggested that a complex system is not needed to manage data standards and an Excel spreadsheet is often sufficient. Though only companies B and D explicitly mentioned the use of CDISC terminology in the answer to Q1, subsequent answers confirmed all four companies were using the CDISC suite of standards. Three of the companies (A, B and C) use SAS (Statistical Analysis System—a statistical software) in addition to Excel for modelling and analysis.

TABLE 1 Summary of the responses from the four companies to the 10 questions about data standards used with their institutions

	Company A	Company B	Company C	Company D
(Q1) Format of data dictionary	Move from Excel to MDR; SAS	Modelled using SAS LSAF, available in Excel, Will be moving to MDR	Excel which are transformed to SAS for further analysis	Excel with CDISC controlled terminology.
(Q2) How are dictionaries used	Used for creating eCRFs—global forms with mandatory and optional items. Permission required for skipping mandatory or adding new items.	Used for creating eCRFs—permission required for adding new items but granted sparingly to maintain standardized forms.	Used for CRFs. For non CDISC items, inbuilt definitions are used and proposed to CDISC.	Used for CRF creation. Permission of governance board required for addition of new items.
(Q3) One or more dictionaries	One standard dictionary; in addition, study specific forms if not standard.	Several dictionaries that vary between studies	One standard dictionary (CDISC standards) and other non-standard dictionaries.	One master dictionary. Studies may use part of the dictionary.
(Q4) Who maintains dictionaries	Cross-functional teams contribute (including FTEs, clinicians, data management, project management, statistics and standards experts).	Each TA dictionary maintained by a team that can include clinicians and statisticians.	Maintained by a team involving clinicians, statisticians and data standards experts.	Data management expert, CDISC expert, TA clinicians, statistician
(Q5) How is consensus reached	Discussions between requester and various teams - cross-functional regarding content and implementation	Discussions between TA experts, governance and cross-functional teams and one expert on the data collection side.	Internal brainstorming to fit the data items while adhering to industry/ regulatory guidelines.	Disagreements on data representation is very rare. Governance board requests are made if the data cannot be represented with existing tools.
(Q6) Is format interoperable	Yes - within the company	Yes. If something specific is developed by a team, a change request is required	Yes, unless a study is using older standards.	Yes, though past studies may differ with respect to terminologies
(Q7) Time for dictionary development	Continuously being developed, CDISC standard elements added if required	Varies depending on TA and dictionary	CDISC terminologies are simple to add. Longer process for non-CDISC items.	An ongoing process and very time consuming.
(Q8) Cost for development/ maintenance	Several FTEs	No answer could be provided.	No answer could be provided.	No answer could be provided.

TABLE 1 (Continued)

	Company A	Company B	Company C	Company D
(Q9) Knowledge gaps and how c4c can aid	Getting external or legacy data from different systems into the CDISC structure	Normal lab ranges difficult to maintain due to dependence on other factors. Development of PUG.	Continued development of PUG	Not aware of any as they are not involved in many paediatric studies
(Q10) Pain points	Modulation of new questionnaires from CDISC and of not yet CDISC modulated topics, easy presentation of medical items to non-standards teams	Immunophenotyping test not available in CDISC, capturing medication review, COVID-specific issues, normal ranges of weights, different answers to same questions	Lab value ranges, items that are only used in paediatrics (e.g., Tanner)	Biological data is a challenge due to variation by age group, PROM questionnaires

Abbreviations: CDISC, Clinical Data Interchange Standards Consortium; COVID, coronavirus disease; CRF, case report forms; eCRF, electronic CRF; FTE, full time employees; LSAF, Life Sciences Analytic Framework; MDR, metadata repository; PUG, paediatric user guide; SAS, Statistical Analysis System; TA, therapeutic area.

All four companies confirmed the use of data dictionaries for the creation of CRFs (Q2). Company A uses global forms with mandatory and optional items. Company B has global standards that are used for most studies. The exceptions are some observational studies that have different requirements. They provide study teams with eCRF implementation guidelines to help them select appropriate data items for their study's eCRF. Company C stated that despite attempting to use CDISC standards they have issues with paediatric studies since several paediatric terms are unavailable in CDISC. Company D has a library of codes which are continuously updated based on authorised requests they receive. They consult CDISC to see if a suitable data item is available in their standards before creating their own.

Adherence to CDISC standards and dealing with data items not available in CDSICs were themes that occurred in multiple questions. All companies had strict processes that needed to be adhered to when adding new items, which may not be available in CDISC. These were relevant themes in Q2 ('dealing with non CDISC data items') and Q6 ('exceptions'). While company C proposed new fields directly to CDISC, the other three required intra-company permission to add/edit fields. Hence, it was apparent that all companies were committed to interoperability ('affirmation' theme in Q6)—at least at the inter-company level. While data acquired from other companies or older datasets could create challenges to interoperability ('exceptions' theme in Q6), the data being generated currently adhere to strict standards.

Three of the companies (A, C and D) maintain one master dictionary (Q3), which aids interoperability. All three the companies have provisions for using more specific versions of the dictionary Companies A and C maintain additional dictionaries for non-CDISC standards terminology. Company D has provisions to use a subset of the dictionary wherever suitable. Company B was the only company that using multiple dictionaries.

The maintenance (Q4) of the dictionaries required multi-disciplinary teams. Several more specific themes regarding expertise emerged from the answers. These included clinicians, statisticians, data managers, governance teams and data standards experts. While the salaries of these employees (at least partially) would be included in the cost for maintaining a dictionary, only company A stated this explicitly. No other information about dictionary maintenance costs was available from any of the companies (Q8).

TABLE 2 Thematic analysis summary from the interviews

	Theme	Number of annotations	Number of companies	Example annotations
(Q1) Format of data dictionary	Excel	10	4	'Excel spreadsheet', 'Excel file', 'Excel form'
	SAS	4	2	'SAS dataset'
	Transition to MDR	7	2	'moving to a new system called MDR'
	CDISC adherence	4	2	'CDISC controlled terminology', 'CDISC file'
(Q2) How are dictionaries used	CRF creation	9	4	'build eCRFs', 'creation of the CRF'
	Dealing with exceptions	5	2	'use their own definitions', 'create new code'
	Dealing with non CDISC data items	3	2	'no CDISC reference', 'consult CDISC'
(Q3) One or more dictionaries	One standard dictionary	3	3	'one standard', 'library of code lists'
	Smaller or more specific dictionaries	3	3	'light version available', 'also have non-standard dictionaries'
(Q4) Who maintains dictionaries	Large team	4	4	'group of people', 'several FTEs', 'by a team'
	Clinicians	7	4	'clinical', 'clinicians', 'TA experts'
	Statisticians	5	4	'statistics', 'statistician'
	Governance	4	2	'governance', 'governance department'
	Data managers	4	3	'data management experts', 'data manager'
	Data standards experts	3	2	'data standards team', 'standards experts'
(Q5) How is consensus reached	Internal discussions	5	3	'would brainstorm within the team', 'discussion with the cross-functional team'
(Q6) Is format interoperable	Affirmation	7	4	'yes', 'don't have multiple dictionaries', 'pool data'
	Exceptions	4	4	'exceptions to the rule', 'past studies may differ'
(Q7) Time for dictionary development	Lengthy and time-consuming process	4	4	'very time-consuming', 'existed for many years and added to over time'
	CDISC adherence	3	2	'item not available in CDISC', 'have to use CDISC standards'
(Q8) Cost for development/maintenance	Unable to answer	3	3	'couldn't answer', 'unable to say', 'unable to advice'
(Q9 and Q10) Knowledge gaps and pain points and how c4c can aid	Lab values (particularly ranges)	8	3	'lab assessments', 'lab value ranges', biological data'
	Data items not being available on CDISC	4	2	'CDISC standards not available for everything', 'CDISC won't add immunophenotyping tests'

TABLE 2 (Continued)

Theme	Number of annotations	Number of companies	Example annotations
Different answers to same question	3	2	'evaluator variability', 'answered by patient – can be very difficult'
Support for PUG	2	2	'could be of interest', 'useful for understanding'

Abbreviations: CDISC, Clinical Data Interchange Standards Consortium; CRF, case report forms; eCRF, electronic CRF; FTE, full time employees; LSAF, Life Sciences Analytic Framework; MDR, metadata repository; SAS, Statistical Analysis System; TA, therapeutic area.

1. What is the format of your Data Dictionary / Dictionaries? (e.g. excel, PDF, modelling tool)

The data dictionaries are currently modelled using the SAS LSAF (life sciences analytics framework) and are available for studies to use in an excel format in SharePoint. They will be moving to a new system called MDR. [redacted] is unsure how the dictionaries will be shared using this new tool but thinks it will likely be in an excel spreadsheet. Avril asked if they have encountered any issues or lost any detail exporting to an Excel spreadsheet. There have been no issues that [redacted] is aware of but could follow-up with members of [redacted] team to check. Rebecca asked if the metadata is searchable within the spreadsheet. [redacted] said not currently. The datasets are represented in different tabs in the spreadsheet and each item shows if it has come from CDISC terminology (from NCI) or the sponsor (defined by [redacted]) terminology. The NCI terminology is updated every six months. The MDR tool should allow you to search the metadata and view therapeutic area (TA) datasets. In the MDR tool, will also allow you to see the controlled terminology for that TA. This will allow the user to see a subset specific to that TA. [redacted] shared [redacted] screen and showed how the dictionaries look both within in their current tool and in Excel form.

FIGURE 1 An example of annotations for the thematic analysis (Question 1 for company B). The annotated phrases refer to the following themes: Red—SAS, Green—MDR, Yellow—Excel, Magenta—CDISC adherence. For follow-up questions, the same annotation appearing in both the question and answer was annotated once. Identifying information of the interviewee and company are redacted

In cases of disagreements in the representation of data items, consensus was reached through discussions between various teams (Q5). These include governance teams (companies A and D), cross-functional teams, medical experts and data collectors (companies A and B). While company C provided a broader answer (internal discussions), these teams may be involved. In addition, company D claimed that disagreements over data representation were very rare. Finally, none of the companies provided a direct answer to the temporal constraints on building a data dictionary (Q7) but all agreed that it was a continuous and time-consuming process. Company D specifically mentioned that they are unable to keep up with the quarterly updates to CDISC terminologies. Term updates are made only if there are multiple requests for the term to be included.

There were two major issues raised by the companies in the knowledge gaps and pain points questions (Q9, 10). The issues were: (1) maintaining normal lab ranges, and (2) dealing with terms not covered in CDISC.

Three companies (B, C and D) raised the issue of maintaining normal lab value ranges (though company D used a more general term—'biological data'). For paediatrics, normal lab ranges are particularly challenging to maintain as they may depend on age, gender, ethnicity, condition etc. This is in addition to the issues standardising across multiple assay methods and approaches to calibration, and, in neonates identifying subjects for reference ranges. While certain suggested reference ranges are available,[28] company B urges against the inclusion of normal ranges and

instead called upon c4c and other expert advice groups to issue common sense guidelines. Incorrect reference ranges can have serious impacts on patient safety as adverse events may remain unrecorded.

The second issue about terminologies not available in CDISC was raised by three companies (A, B and C). For paediatric studies, this is further compounded due to therapeutic area user guides not being available for most diseases. Due to existing regulations all companies must use CDISC standards. As mentioned above, each company has its own process for including non-CDISC terminology. Companies B and C agreed that a PUG developed by c4c and CDISC would have substantial added value in attempting to fill this gap, by covering a broad spectrum of paediatric conditions. Company C remarked that if c4c provided a data modelling and standardisation resource, this could be of interest to them. Company A provided a more general response about how their involvement in c4c has improved their understanding of how academia works.

Companies B and D raised the issue of how the same answer can be expressed in different ways. For example, company B mentioned the answer provided by different guardians (parents, grandparents, stepparents) can be different. Company D had similar issues with patient reported outcome measure (PROM) questionnaires. Some of the other issues raised by the companies were easy presentation of medical items to non-standards teams (company A), capturing medication review, addressing COVID-specific challenges, and normal ranges for weights (company B).

4 | DISCUSSION

The feedback from the interviews highlighted the need for paediatric-specific tools. Standards developed by CDISC are well understood by the pharmaceutical industry, and their value is reflected in the requirement for their use in electronic regulatory submissions worldwide. However, CDISC standards currently lack the paediatric specificity required by the sponsors of clinical trials. CDISC standards are also not well understood in academic clinical research. Based on these interviews, c4c is working on a four-point plan to enable standardisation of paediatric data items: (a) a cross-cutting paediatric data dictionary (CCPDD) for disease-agnostic data items along with a clinical modelling tool (CMT) for better visualisation; (2) the previously mentioned PUG that extends to therapeutic area standards, some disease-specific metadata, examples, and guidance on implementing CDISC standards for a variety of uses, including global regulatory submission; (3) Extension to disease-specific data items, and (4) Extension to real-world data. These tasks precisely align with the knowledge gaps and pain points identified through these interviews. Moreover, these tools are being developed in close collaboration with CDISC as the companies must use CDISC approved standards due to existing regulations.

The CCPDD partially address the knowledge gap of dealing with non CDISC data items. It provides guidelines on the representation of disease agnostic data items that are commonly used in paediatrics with the intention of generating interoperable data with potential for reuse. The first version took 16 months to develop and was successfully used to manually create CRFs in three proof of viability (PoV) trials—A New Posaconazole Dosing Regimen for Paediatric Patients With Cystic Fibrosis and Aspergillus Infection (cASPerCF, EudraCT Number: 2019-004511-31), Kawasaki Disease Coronary Artery Aneurysm Prevention trial (KD-CAAP, EudraCT Number: 2019-004433-17) and Prophylactic Treatment of the Ductus Arteriosus in Preterm Infants by Acetaminophen (TREOCAPA - EudraCT Number: 2019-004297-26). The CCPDD acts as a 'look up table' listing paediatric data items that are commonly collected in clinical trials. The items are grouped into four areas: demographics, vital signs, pubertal status, and others. The data dictionary provides definitions of each item as well as suggested clinical domains (e.g., vital signs), qualifiers (such as subject position), units (e.g., kg), and CDISC standards referenced wherever available. A drawback of this first version is the development in Microsoft Excel that led to a flat structure with poor visualisation. Hence, c4c has piloted a CMT.[29] The CMT allows data items to be viewed as HTML documents, a tree structure, or be exported in various machine-readable formats (e.g., XML).

The PUG extends the CCPDD to tackle the value-range and questionnaire issues raised during the interviews. This PUG development process started by convening an online meeting last year between paediatricians, statisticians,

c4c working group members, CDISC data standards experts and representatives from academia and industry. These experts followed the CDISC standards development process, initially working through the scoping stage to identify the most relevant concepts for paediatric studies for inclusion in the PUG. This has included the identification of new paediatric specific concepts such as 'Birth Complications', and new Questionnaires, Ratings and Scales (QRS) instruments including the Bayley Scales of Infant and Toddler Development and the APGAR score. Value ranges or guidelines are provided wherever applicable. The PUG is mostly complete and currently available for public review.[30] Once finalised (December 2022), the PUG will be freely available on the CDISC website and can be used to represent data in paediatric studies to expedite the regulatory review process, reduce time to market, and drive operational efficiencies within organisations that use them.

Extensions to disease-specific paediatric data items and real-world data are at their initial stages. Therapeutic area standards extend the CDISC foundational standards to represent data that pertains to specific disease areas. To address the identified knowledge gaps, the solutions must go beyond cross-cutting items. A kick-off workshop for the disease-specific extension was held in Rome in April 2022 with the first focus area being metabolic diseases. During the workshop it was also decided to leverage electronic health records (EHR) standards such as OMOP or Fast Healthcare Interoperability Resources (FHIR) through transformations with CDISC. A FHIR to CDISC transformation guide was jointly released last year,[31] while transformations with OMOP have also been tested in the literature.[32] While such transformations can substantially reduce workloads, they are generally accompanied by some loss of information.

The extension to real-world data kicked off in June 2022 and has been split into two strands—(a) real-world data as a comparator arm, and (b) real-world data for post-marketing surveillance. The transformations discussed for the EHR standards would be highly relevant here.

Our study has certain limitations. The number of companies interviewed were small. While each question had at least one theme that included three companies (Table 2), a larger or different set of companies may have led to different conclusions. On the other hand, the study involved a large manual component. Hence, expanding it to more companies would come at significant cost. The interviewees of all four companies, were based in Europe. It is unclear how the answers would generalise to the rest of the world.

5 | CONCLUSIONS

By facilitating direct engagement between academic and industry experts c4c has identified the knowledge gaps pertinent in paediatric data standards. With the development of the CCPDD and PUG in collaboration with CDISC and with planned extensions to disease-specific and real-world data, c4c hopes to promote standardisation of paediatric data items and enable an easier process for paediatric clinical trials.

ACKNOWLEDGEMENTS

The authors would like to thank all interviewees for their participation. This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement no. 777389. The Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

CONFLICTS OF INTEREST

No disclosures.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ETHICS STATEMENT

No human subjects were used in this study.

DISCLAIMER

The publication reflects the authors' view and neither IMI nor the European Union, EFPIA, or any Associated Partners are responsible for any use that may be made of the information contained therein.

ORCID

Anando Sen https://orcid.org/0000-0001-6761-2632
Avril Palmeri https://orcid.org/0000-0002-6171-5793
Joanne Lee https://orcid.org/0000-0001-5359-2941
Victoria Hedley https://orcid.org/0000-0002-4437-2469
Rebecca Leary https://orcid.org/0000-0002-8039-5334
Volker Straub https://orcid.org/0000-0001-9046-3540

REFERENCES

1. European Medicines Agency and its Paediatric Committee. 2017 Commission Report on the Paediatric Regulation. Updated 2016; 2022. Accessed February 1, 2022. https://ec.europa.eu/health/medicinal-products/medicines-children/2017-commission-report-paediatric-regulation_en
2. Hwang TJ, Tomasi PA, Bourgeois FT. Delays in completion and results reporting of clinical trials under the Paediatric Regulation in the European Union: a cohort study. *PLoS Med*. 2018;15(3):e1002520. https://doi.org/10.1371/journal.pmed.1002520
3. Lim A, Cranswick N. Clinical trials, ethical issues and patient recruitment: an australian perspective. *Paediatr Perinat Drug Ther*. 2003;5(4):183-187. https://doi.org/10.1185/146300903774115775
4. Caldwell PH, Murphy SB, Butow PN, Craig JC. Clinical trials in children. *Lancet*. 2004;364(9436):803-811. https://doi.org/10.1016/s0140-6736(04)16942-0
5. von Baeyer CL. Children's self-reports of pain intensity: scale selection, limitations and interpretation. *Pain Res Manag*. 2006;11(3):157-162. https://doi.org/10.1155/2006/197616
6. Green DJ, Burnham JM, Schuette P, et al. Primary endpoints in pediatric efficacy trials submitted to the US FDA. *J Clin Pharmacol*. 2018;58(7):885-890. https://doi.org/10.1002/jcph.1109
7. Joseph PD, Craig JC, Caldwell PHY. Clinical trials in children. *Br J Clin Pharmacol*. 2015;79(3):357-369. https://doi.org/10.1111/bcp.12305
8. Zurynski Y, Gonzalez A, Deverell M, et al. Rare disease: a national survey of paediatricians' experiences and needs. *BMJ Paediatr Open*. 2017;1(1):e000172. https://doi.org/10.1136/bmjpo-2017-000172
9. Bavisetty S, Grody WW, Yazdani S. Emergence of pediatric rare diseases: review of present policies and opportunities for improvement. *Rare Dis*. 2013;1:e23579. https://doi.org/10.4161/rdis.23579
10. Phillips B, Davies HT, Preston J, Stones SR. Framework to help design and review research involving children. *Arch Dis Child*. 2019;104(6):601-604. https://doi.org/10.1136/archdischild-2018-315119
11. Maagdenberg H, Vijverberg SJ, Bierings MB, et al. Pharmacogenomics in pediatric patients: towards personalized medicine. *Paediatr Drugs*. 2016;18(4):251-260. https://doi.org/10.1007/s40272-016-0176-2
12. Davies HT, Phillips B, Preston J, Stones SR. Making research central to good paediatric practice. *Arch Dis Child*. 2019;104(4):385-388. https://doi.org/10.1136/archdischild-2018-315117
13. conect4children is a pan-European clinical trial network. Accessed February 2, 2022. https://conect4children.org/
14. Turner MA, Hildebrand H, Fernandes RM, et al. The conect4children (c4c) consortium: potential for improving European clinical research into medicines for children. *Pharm Med*. 2021;35(2):71-79. https://doi.org/10.1007/s40290-020-00373-6
15. Rasmussen K, Bero L, Redberg R, Gøtzsche PC, Lundh A. Collaboration between academics and industry in clinical trials: cross sectional study of publications and survey of lead academic authors. *BMJ*. 2018;363:k3654. https://doi.org/10.1136/bmj.k3654
16. Banks M. Biopharma-academic collaborations in 2021. *Appl Clin Trials*. 2021;30(11).

17. Davie N. *Academia-industry Collaboration in Translational Medicine*. University of Oxford; 2016.
18. Sandercock P. Negative results: why do they need to be published? *Int J Stroke*. 2012;7(1):32-33. https://doi.org/10.1111/j.1747-4949.2011.00723.x
19. Nimpf S, Keays DA. Why (and how) we should publish negative data. *EMBO Rep*. 2020;21(1):e49775. https://doi.org/10.15252/embr.201949775
20. Blacketer C. The common data model. In: *The Book of OHDSI*. Observational Health Data Sciences and Informatics; 2019.
21. Tamm K. How and when to use APIs: CDISC ODM XML. In: *OpenClinica Reference Guide*. OpenClinica; 2020.
22. FDA. Study Data Standard: What You Need to Know. Accessed February 9, 2022. https://www.fda.gov/files/drugs/published/Study-Data-Standards--What-You-Need-to-Know.pdf
23. Bobbitt DR, LeRoy B, Palmer A, et al. Clinical data standards and the new world of research science, technology, and data sources. *Journal of the National Institute of Public Health*. 2019;68(3):194-201. https://doi.org/10.20683/jniph.68.3_194
24. Facile R, Muhlbradt EE, Gong M, et al. Use of clinical data Interchange standards consortium (CDISC) standards for real-world data: expert perspectives from a qualitative delphi survey. *JMIR Med Inf*. 2022;10(1):e30363. https://doi.org/10.2196/30363
25. Utidjian LH. *Identifying and Understanding Data Quality Issues in a Pediatric Distributed Research Network*. American Academy of Pediatrics; 2015.
26. Ranade-Kharkar P, Narus SP, Anderson GL, Conway T, Del Fiol G. Data standards for interoperability of care team information to support care coordination of complex pediatric patients. *J Biomed Inf*. 2018;85:1-9. https://doi.org/10.1016/j.jbi.2018.07.009
27. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101. https://doi.org/10.1191/1478088706qp063oa
28. Klarmann D, Hintereder G, Thomas L. *Reference Intervals and Orienting Ranges for Clinical Chemistry and Laboratory Medicine in Fetuses, Newborn and Adolescents*. Clinical Laboratory Diagnostics 2020; 2021. Accessed April 5, 2022. https://www.clinical-laboratory-diagnostics-2020.com/k54.html
29. Amadi C, Leary R, Palmeri A, et al. How can a clinical data modelling tool Be used to represent data items of relevance to paediatric clinical trials? learning from the Conect4children (c4c) consortium. *Appl Sci*. 2022;12(3):1604. https://doi.org/10.3390/app12031604
30. CDISC. Pediatrics User Guide v1.0. CDISC; 2022. Accessed August 12, 2022. https://www.cdisc.org/standards/therapeutic-areas/pediatrics/pediatrics-user-guide-v1-0
31. CDISC. FHIR to CDISC Joint Mapping Implementation Guide v1.0; 2022. Accessed August 12, 2022. https://www.cdisc.org/standards/real-world-data/fhir-cdisc-joint-mapping-implementation-guide-v1-0
32. Orlova MKADA, Ong E-MDR, Biedermann P, Wetheril G, Klebanov G. Pathways for Advanced Transformation of CDISC SDTM Data Sets into OMOP CDM; 2020.

AUTHOR BIOGRAPHIES

Anando Sen is a Data Scientist at the John Walton Muscular Dystrophy Research Centre, Newcastle University working on the conect4children project. He has over 10 years of experience in mathematical and statistical modelling for various healthcare applications. His research interests include medical imaging, clinical trials, and data analysis

Avril Palmeri is the Data Officer on the conect4children. She is based at the John Walton Muscular Dystrophy Research Centre at Newcastle University. Previously, she was involved in Share4Rare project (online platform for neuromuscular diseases patients) and RD-Connect project (rare diseases analysis).

Joanne Lee is an Education Coordinator at John Walton Muscular Dystrophy Research Centre at Newcastle University. She works on the conect4children project (development of short courses for clinical trials unit staff) and the European Joint Programme for Rare Diseases project (development of the Advisory Committee for Therapeutics toolkit)..

Victoria Hedley is the Rare Disease Policy Manager and Co-Lead for the Newcastle University Centre of Research Excellence in Rare Disease. She is based at the John Walton Muscular Dystrophy Research Centre. Victoria has

led the Rare 2030 foresight project (knowledge-based activities), RD-ACTION (policy), the EU Joint Action for RD and conect4children (data-related activities).

Jacques Thuet is Data Management & CDISC Expert at Institut de Recherches Internationales Servier. He has over 10 years of experience in Data Management and CDISC Standard and is currently in charge of the Study Data Tabulation Model production management.

Perrine Lignon, Head of Clinical data management at Servier. With 20 years' experience in the fields of data management, business organisation and transformation, she is passionate by the evolution of drug development and decentralisation of clinical trials. She has significant international experience having worked with teams in different countries.

Valerie Cotonnec is a Senior Data Management and Central Monitoring Expert at Servier. With 21 years' experience in the data management, she has managed both projects and processes. She has developed a great interest for the risk-based management and central monitoring activities.

Rebecca Leary is a Senior Project Manager managing the conect4children project. Rebecca is based at the John Walton Muscular Dystrophy Research Centre, Newcastle University. Previously, she managed the neuromuscular European Reference Network, and was part of the global TREAT-NMD alliance, which facilitates translational research (therapy for patients with rare neuromuscular diseases).

Sinéad Nally (Nic An Fhailí) is a Clinical Development Director (Neuroscience) and a member of the Paediatric Centre of Excellence, Novartis. Sinéad has a long-standing interest in paediatric clinical research, data harmonisation and data interoperability.

Volker Straub is Professor of Neuromuscular Genetics, Deputy Dean of the Newcastle University Institute of Translational and Clinical Research, and Director of the university's John Walton Muscular Dystrophy Research Centre. He has a long-standing interest in the pathogenesis of genetic muscle diseases, with a focus on translational research in muscular dystrophies.

How to cite this article: Sen A, Palmeri A, Lee J, et al. Understanding paediatric data standards challenges through academia-industry partnerships: a conect4children (c4c) qualitative study. *Int J Health Plann Mgmt*. 2022;1-14. https://doi.org/10.1002/hpm.3592