Journal of Cystic Fibrosis xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Cystic Fibrosis

journal homepage: www.elsevier.com/locate/jcf

Original Article

Data accuracy, consistency and completeness of the national Swiss cystic fibrosis patient registry: Lessons from an ECFSPR data quality project

Lara Wolf [a,b], Jakob Usemann [a,c], Eugénie Collaud [a], Marie-France Derkenne [d], Reta Fischer [e], Maxime Hensen [f], Michael Hitzler [g], Markus Hofer [b], Demet Inci [a], Sarosh Irani [h], Kathleen Jahn [i], Angela Koutsokera [d], Rachel Kusche [h,j], Thomas Kurowski [k], Philipp Latzin [l], Dagmar Lin [m], Laurence Mioranza [n], Alexander Moeller [a], Anne Mornand [o], Dominik Mueller-Suter [j], Christian Murer [p], Lutz Naehrlich [q,u], Jérôme Plojoux [r], Nicolas Regamey [g], Romy Rodriguez [l], Isabelle Rochat [n], Alain Sauty [s], Macé Schuurmans [k], Michaela Semmler [m], Daniel Trachsel [c], Anna-Lena Walter [t], Andreas Jung [a,b,u,*]

[a] Department of Respiratory Medicine, University Children's Hospital Zurich, Switzerland
[b] Departments of Pulmonology and Paediatric Pulmonology, Cantonal Hospital Winterthur, Switzerland
[c] Department of Pulmonology, University Children's Hospital Basel, Switzerland
[d] Department of Pulmonology, Lausanne University Hospital, Switzerland
[e] Quartier Bleu, Pulmonology Outpatient Clinic, Berne, Switzerland
[f] Department of Paediatrics, Cantonal Hospital Fribourg, Switzerland
[g] Department of Paediatric Pulmonology, Cantonal Hospital Lucerne, Switzerland
[h] Department of Pulmonology, Cantonal Hospital Aarau, Switzerland
[i] Clinic of Respiratory Medicine and Pulmonary Cell Research, University Hospital Basel, Switzerland
[j] Department of Paediatric Pulmonology, Cantonal Hospital Aarau, Switzerland
[k] Department of Pulmonology, University Hospital Zurich, Switzerland
[l] Division of Paediatric Respiratory Medicine and Allergology, Department of Paediatrics, Inselspital Berne University Hospital, Switzerland
[m] Department of Pulmonology, Inselspital Berne University Hospital, Switzerland
[n] Paediatric Pulmonology and Cystic Fibrosis Unit, Division of Paediatrics, Department Woman-Mother-Child, Lausanne University Hospital, Switzerland
[o] Department of Paediatric Pulmonology, Geneva University Hospital, Switzerland
[p] Department of Pulmonology, Cantonal Hospital Lucerne, Switzerland
[q] Department of Paediatrics, Justus-Liebig-University Giessen, Germany
[r] Deparment of Pulmonology, Geneva University Hospital, Switzerland
[s] Department of Pulmonology, Réseau Hospitalier Neuchâtelois, Switzerland
[t] Department of Pulmonology, Cantonal Hospital St. Gallen, Switzerland
[u] European Cystic Fibrosis Society Patient Registry, Karup, Denmark

ARTICLE INFO

Keywords:
Patient registries
Data quality
Data quality audit
Cystic fibrosis

ABSTRACT

Background: Good data quality is essential when rare disease registries are used as a data source for pharmacovigilance studies. This study investigated data quality of the Swiss cystic fibrosis (CF) registry in the frame of a European Cystic Fibrosis Society Patient Registry (ECFSPR) project aiming to implement measures to increase data reliability for registry-based research.

Methods: All 20 pediatric and adult Swiss CF centers participated in a data quality audit between 2018 and 2020, and in a re-audit in 2022. Accuracy, consistency and completeness of variables and definitions were evaluated, and missing source data and informed consents (ICs) were assessed.

Results: The first audit included 601 out of 997 Swiss people with CF (60.3 %). Data quality, as defined by data correctness ≥95 %, was high for most of the variables. Inconsistencies of specific variables were observed

Abbreviations: CF, cystic fibrosis; CFTR, Cystic fibrosis transmembrane conductance regulator; ECFS, European Cystic Fibrosis Society; ECFSPR, European Cystic Fibrosis Society Patient Registry; FEV1, Forced expiratory volume in one second; IC, Informed consent; MRSA, Methicillin-resistant Staphylococcus aureus; PAES, Post-authorization efficacy study; PASS, Post-authorization safety study.

* Corresponding author at: Department of Respiratory Medicine, University Children's Hospital Zurich, Steinwiesstrasse 75, 8032 Zurich, Switzerland.
E-mail address: andreas.jung@kispi.uzh.ch (A. Jung).

https://doi.org/10.1016/j.jcf.2023.08.015
Received 2 May 2023; Received in revised form 3 August 2023; Accepted 31 August 2023

because of an incorrect application of the variable definition. The proportion of missing data was low with <5 % for almost all variables. A considerable number of missing source data occurred for CFTR variants. Availability of ICs varied largely between centers (10 centers had >5 % of missing documents). After providing feedback to the centers, availability of genetic source data and ICs improved.

Conclusions: Data audits demonstrated an overall good data quality in the Swiss CF registry. Specific measures such as support of the participating sites, training of data managers and centralized data collection should be implemented in rare disease registries to optimize data quality and provide robust data for registry-based scientific research.

1. Introduction

Cystic fibrosis (CF) is a rare inherited disease with an incidence of 1:3357 newborns in Switzerland [1]. In the era of CFTR modulator treatment, pharmacovigilance studies have shifted to confirming results of phase-III studies in real-life scenarios, and patient registries have been recognized as data platforms for real-world evidence. Optimal data quality is a requirement for delivering robust and trustworthy clinical information in this context. This is especially important for rare diseases, as registries can close the gap of lacking large clinical trials due to low patient numbers. Moreover, rare disease registries can be used as platforms to specifically identify and recruit patients for randomized controlled clinical trials [2]. Consequently, there is a worldwide increasing interest in research on data quality within patient registries [2–6].

The main components of data quality are accuracy (registry data matching the medical records), consistency (registry data matching the variable definitions), and completeness (missing registry information that is present in clinical source data). Only sufficient data quality in these three domains qualifies a registry to be used as a robust data platform for clinical and epidemiological research. Criteria for good practice and evaluation for patient registries have been suggested in the past. These were internationally acknowledged and consist of a framework requirement of data quality, research quality, and evidence quality [3]. Other groups have also published recommendations for improving the quality of rare disease registries, including aspects of governance, infrastructure, documentation, training, and quality audit, amongst others [4]. Since the implementation and sustainability of quality criteria is sometimes difficult to achieve, especially in large registries, surveillance tools such as data quality audits have been established. These audits are based on internal or external independent and systematic examinations of data conformity to predetermined standards or criteria. The findings - usually resulting from onsite visits - are subsequently reported to all staff members, indicating the areas for improvement and highlighting good practice of data collection.

The European Cystic Fibrosis Society Patient Registry (ECFSPR) has undertaken large efforts to enhance data quality, as well as improve data collection coverage within Europe [7–9]. As a result, the European Medicines Agency accepted the ECFSPR as a data source for post-authorization efficacy (PAES) and safety (PASS) studies, following a qualification opinion [10]. In 2018, the ECFSPR started data quality audits in its member countries including Switzerland, focusing on several larger CF centers in each country [8]. This was to ensure Europe-wide quality standards in response to increasing interest in registry-based research, including pharmacovigilance studies. In the ECFSPR annual report 2020, the first summarizing results of these onsite visits have been published, demonstrating an overall good data completeness and identifying specific areas of future improvement [9]. Switzerland contributes data to the ECFSPR since 2008. The Swiss CF registry has launched a national data quality program between 2020 and 2022 following the ECFSPR data audits, with onsite-visits before and after feedback and training of the CF centers. The aim of this project was to analyze data quality in a data quality audit to assess accuracy, consistency, and completeness of variables and definitions, to define shortcomings and difficulties in accomplishing good data quality, and to provide recommendations on how to optimize data collection in larger patient registries.

2. Methods

2.1. Patient registry

The ECFSPR collects demographic and clinical data from people with CF (pwCF) in Europe. Subjects are included according to defined regulations after signing an informed consent [11]. Annual data collection includes information on demographics, diagnosis, cystic fibrosis transmembrane conductance regulator (CFTR) variants, growth parameters, lung function, microbiology, treatment, complications, organ transplantation, and mortality. In the ECFSPR, data from more than 52,000 pwCF are included from 39 participating countries [9]. The Swiss national CF patient registry participates in the continuous annual ECFSPR data collection, using a joint web-based data collection platform called ECFSTracker. All 20 Swiss CF centers participate in the registry, resulting in a coverage of >98 % of all Swiss children and adults with CF.

2.2. Data quality audits

With regards to the ECFSPR data audit, the first onsite visits were conducted in 2018 by a team from the ECFSPR and the Interdisciplinary Centre for Clinical Trials, University Medical Centre Mainz, Germany. Annual data from the 2016 data collection were verified at source level in four countries (Austria, Portugal, Slovakia, Switzerland). During this first period of onsite visits, four out of 20 pediatric and adult Swiss CF centers were included (**Table S1**). To complete data validation on a nationwide level in Switzerland, a national auditor conducted onsite visits from February to August 2020 and verified annual data from the 2018 data collection at source level in the 16 remaining Swiss CF centers (**Table S1**). Data structure was identical between 2016 and 2018. The ECFSPR variable definitions also remained unchanged expect for liver disease, as variceal bleeding was added to the definition in 2018 (**Table S2**). Therefore, comparison between the two years of data collection seemed appropriate. Subjects in each center were randomly selected for data validation. In each center, as many individuals as possible were included for validation in the time frame of one working day. Only pwCF with existing written informed consent (IC) could be included in the data validation. After feedback of the audit results and training of the CF centers, we conducted a scheduled re-audit in 2022 (2021 data collection) on the major shortcomings of the first audits, specifically focusing on the availability of genetic source data and ICs. One center could not be re-audit for organizational reasons (Table S1). A summary of the different audit phases is displayed in Table 1.

2.3. Variables and definitions

The selection of variables focused on items of particular importance for key registry reporting in the ECFSPR, including variables that have been highlighted as challenging by users. Those included demographic data, diagnostic and transplant data, anthropometric data, best lung function measurement, results of selected bacterial infections, medications, and complications (Table 2 and **S2**). In 2020, additional variables

Table 1
Characteristics of the three phases of data quality audits between 2018 and 2022.

Audit year	Data year	Audit by	Centers included*	Aims
2018	2016	ECFSPR	4 largest CF centers (2 pediatric, 2 adult)	Data validation of key registry variables**
2020	2018	Swiss CF registry	All remaining 16 CF centers	Completion of data validation on a nationwide level
2022	2021	Swiss CF registry	19 CF centers	Re-evaluation of availability of genetic source data and informed consents

* for details, see Table S1.
** according to an ECFSPR audit protocol, focusing on variables of importance for registry-based pharmacovigilance studies.

for bacterial colonization and medications including CFTR modulator therapies that were not yet included in the 2018 audit were evaluated for the 16 centers audited. Definitions of variables were used according to ECFSPR (**Table S2**). In addition to data quality and completeness, availability of written informed consents to participate in the ECFSPR was evaluated.

Correct data reflects accuracy (values entered in ECFSTracker matching the medical record), consistency (values entered in ECFSTracker matching the definitions set by ECFPR), and completeness (missing information in ECFSPR that are present in clinical source data or missing source data or missing ICs, respectively). High data quality was defined as ≥95 % correctness for all items.

2.4. Ethics approval and informed consent

This research project complies with the principles of the Declaration of Helsinki. The study has been approved by the cantonal ethics committees of all participating centers. Written informed consent is required from all pwCF included in the Swiss CF registry data collection.

2.5. Statistical analysis

Data quality for accuracy, consistency and completeness was defined as good if correctness was ≥95 %, and poor if >5 % were incorrect, according to ECFSPR definitions ([8,9]). Data quality was compared across the study centers according to center size defined by the number of individuals (small n<35, medium n=35-74, large n>=75) using the Kruskall-Wallis test. Cut-offs of number of individuals per center reflect the distribution of center size in Switzerland, with four large centers, eight medium-sized centers and eight smaller centers. Data quality was also compared between adult and pediatric CF centers as well as according to the professional background of the local data managers using the Mann-Whitney test. Descriptive statistics and non-parametric tests were performed using SPSS 25 (IBM, Illinois/Chicago, USA). Statistical significance was defined as p-value <0.05, and two-sided tests were used.

3. Results

In the first onsite visits in 2018 and 2020, 601 out of 997 individuals with existing registry records (60.3 %) were audited (25.3-100 % per center, depending on patient number and organization of clinical records in the centers) (**Table S3**). Data quality, as defined by correctness ≥95 %, was high for most of the variables (Table 2). Incorrect data of >5 % were observed for body height and weight, FEV1, and liver disease. While in the ECFSPR FEV1 is defined as the best value during the year of follow-up, the last measurement in the calendar year was used by mistake in some cases. Likewise, body weight and height should be collected at the date of best FEV1, not at the last clinical visit during the year of follow-up. Due to lack of a precise ECFSPR definition for liver disease and hemoptysis, data correctness was judged to be debatable in a low number of individuals. The proportion of missing data was very low and did not exceed 5 % in almost all variables. A considerable number of missing source data occurred for CFTR variants only.

A high variability between the centers for the availability of written informed consent to participate in the ECFSPR was found (Table 3). The proportion of missing ICs ranged from 0 and 69.2 %, and in 10 centers >5 % of ICs were missing. The overall number of missing ICs for Switzerland was 15.5 % in the 2018/2020 data quality audits. There was

Table 2
Data quality (n records, %) for all variables included in the data audit. Data inaccuracy in >5 % of records is depicted in bold. Microbiology data refer to oropharyngeal swab, sputum, or bronchoalveolar lavage fluid.

Variable	Correct	Incorrect	Missing entry	No source data	Debatable	Total
Gender	469 (99.8)	1 (0.2)	0	0	0	470
Date of birth	468 (99.6)	2 (0.4)	0	0	0	470
CFTR variant 1	311 (66.5)	2 (0.4)	0	**155 (33.1)**	0	468
CFTR variant 2	303 (65.0)	4 (0.9)	3 (0.6)	**156 (33.5)**	0	466
CFTR variants both	300 (64.5)	5 (1.1)	3 (0.6)	**157 (33.8)**	0	465
Lung/liver transplant	467 (100)	0	0	0	0	467
Height	402 (86.6)	**42 (9.1)**	9 (1.9)	11 (2.4)	0	464
Weight	375 (81.0)	**71 (15.3)**	9 (1.9)	8 (1.7)	0	463
FEV1	352 (80.5)	**72 (16.5)**	5 (1.1)	8 (1.8)	0	437
Inhaled antibiotics	439 (94.8)	13 (2.8)	4 (0.9)	7 (1.5)	0	463
Inhaled rhDNAse	440 (95.0)	12 (2.6)	2 (0.4)	9 (1.9)	0	463
Hypertonic saline*	268 (95.7)	5 (1.8)	2 (0.7)	5 (1.8)	0	280
Pancreatic enzymes	447 (96.5)	7 (1.5)	3 (0.6)	6 (1.3)	0	463
Insulin therapy	446 (96.3)	7 (1.5)	2 (0.4)	8 (1.7)	0	463
Azithromycin*	262 (93.2)	10 (3.6)	5 (1.8)	4 (1.4)	0	281
CFTR modulators*	272 (96.8)	2 (0.7)	3 (1.1)	4 (1.4)	0	281
Hemoptysis	438 (94.6)	13 (2.8)	2 (0.4)	5 (1.1)	5 (1.1)	463
Liver disease	400 (86.4)	**37 (8.0)**	4 (0.9)	5 (1.1)	17 (3.7)	463
P. aeruginosa	434 (96.0)	11 (2.4)	3 (0.7)	4 (0.9)	0	452
*H. influenzae**	258 (95.2)	8 (3.0)	4 (1.5)	1 (0.4)	0	271
*Achromobacter spp.**	265 (98.1)	1 (0.4)	3 (1.1)	1 (0.4)	0	270
*S. maltophilia**	259 (95.6)	7 (2.6)	4 (1.5)	1 (0.4)	0	271
Burholderia spp.	442 (97.8)	3 (0.7)	3 (0.7)	4 (0.9)	0	452
*MRSA**	265 (98.1)	0	2 (1.1)	3 (0.7)	0	270

* variables added in the 2020 data quality audit.

Table 3
Proportion of missing informed consents per center for the first (years 2018 and 2020) and second (year 2022) data quality audit. Center numbers were defined according to the number of pwCF in ascending order. One center was not audited in 2022 (n/a). Note that the number of audited subjects was different for both time points (60.3 % in 2018/2020 versus 100 % in 2022).

Center no.	ICs missing 2018/2020 audit n (%)	ICs missing 2022 audit n (%)
1	4 (66.7)	n/a
2	0 (0)	1 (11.1)
3	2 (10.0)	0 (0)
4	0 (0)	5 (22.7)
5	4 (17.4)	5 (15.6)
6	0 (0)	0 (0)
7	17 (58.6)	3 (9.7)
8	5 (14.7)	3 (8.1)
9	27 (69.2)	16 (38.1)
10	2 (5.1)	0 (0)
11	2 (4.7)	3 (7.0)
12	10 (21.7)	0 (0)
13	0 (0)	0 (0)
14	1 (2.1)	16 (31.4)
15	1 (2.0)	0 (0)
16	0 (0)	2 (3.5)
17	1 (1.8)	0 (0)
18	18 (25.7)	7 (8.3)
19	4 (8.0)	7 (5.6)
20	1 (1.4)	1 (0.7)
Switzerland	**15.5%**	**6.6%**

no significant difference in the proportion of missing ICs between pediatric and adult CF centers (data not shown).

Since the number of missing genetic source data and ICs were the two most critical findings in the first ECFSPR and Swiss data quality audits in 2018/2020, we performed a re-audit in 2022 specifically for the completeness of these two items in the CF registry. All 1045 records available in the Swiss CF registry were included at that time, except for one center with only 6 subjects registered. The validation visits occurred after all centers had been given feedback on the outcome of the first audit and training on data quality. In this second audit, a total of 28.0 % of genetic source data were still missing (overall improvement of 5.8 % compared to 33.8 % in 2018/2020; data not shown), whereas the overall number of missing ICs was reduced to 6.6 % (improvement of 8.9 %; Table 3). However, the number of centers with >5 % of missing ICs remained unchanged (n=10).

4. Discussion

This study is the first to report outcomes and measures of a national data quality project within the ECFSPR framework. Data audits demonstrated an overall good data accuracy, consistency and completeness in the Swiss CF registry as a result of a long-standing effort to implement provisions to improve data quality in the ECFSPR. These traditionally include written data coding documents on variables and their definitions, standard operating procedures and training sessions on both data collection and software, using built-in error recognition in the software and a thorough statistical error correction and data cleaning procedure during each year of follow-up. Major shortcomings regarding the completeness of genetic source data and ICs could be improved after providing feedback and training to participating centers.

The Swiss national results on data quality were comparable with the summarized findings of all countries participating in the ECFSPR and demonstrated an overall good data quality with most variables achieving data correctness of ≥95 % [9]. Incorrect data in >5 % of the audited records were equally found for height, weight and FEV1 in Europe and Switzerland. The reason for this finding was an inconsistent application of the variable definitions by local data managers. Liver disease also yielded >5 % of incorrect data according to the ECFSPR variable definition in both Europe and Switzerland, and in some cases, correct application of the definition was judged to be debatable. This observation reflects a general haziness in the clinical definition of liver disease and the challenge of translating it into a registry variable. The ECFSPR reacted to this after the first data quality audit and specified the variable definition by adding "variceal bleeding" to the options to choose from for the 2018 data collection. Still, we found persisting numbers of incorrect data in the 2020 audit, reflecting a general uncertainty in clinical diagnosis and documentation of CF liver disease. Similarly, a higher number of incorrect and debatable results were observed for hemoptysis due to a shortcoming of variable definition. Consequently, the ECFSPR adapted the definition for hemoptysis in 2021 to "major volume of blood expectorate >250ml *in a day* during the year of follow-up". These examples demonstrate the need for precise variable definitions that reflect clinical classifications and, at the same time, consider the commonly used terms of medical record documentation.

Few studies have assessed data quality in CF registries, and difficulties in correctly applying variable definitions have been in the focus ever since. An early French study compared clinical source data to registry data in 242 records collected from different CF centers and, similar to our findings, observed a high discrepancy especially for FEV1 data and the associated anthropometric measures as a result of wrongly selected lung function measurements [12]. This study also found considerable deviation of lung function values by incorrect application of spirometry reference standards – a problem the ECFSPR has overcome by collection of lung volumes in liters only and computing predicted values by a standardized reference internally within the ECFSTracker software, applying the Global Lung Function Initiative equations [13]. Issues related to interpretation of variable definitions were also reported by other disease registries, highlighting the need for extensive and repetitive training of local data managers in charge of registry data collection [14].

When comparing European to Swiss audit outcomes, a considerable difference was found for the completeness of genetic source data, with overall 99 % complete data Europe-wide compared to 66.2 % in Switzerland. Upon further investigation in the centers, this was found to be because the genetic lab results were not transferred from the pediatric to the adult center during the transition process. In the era of precision medicine with CFTR modulators, this finding bears the potential risk that, if genetic source data cannot be verified by the prescribing physician, a mutation-specific drug may be incorrectly prescribed. After the CF centers received feedback on the audit results and training on data quality, the proportion of missing genetic lab results improved, however, more than a quarter were still missing indicating that considerable effort needs to be applied to further optimize the availability of clinical source data.

In recent years, data quality programs including audits have been implemented in several national disease registries, with different approaches. The U.S. National Cardiovascular Data Registry assesses data accuracy and completeness, however, datasets failing the validity and completeness checks are rejected and returned to the sites for improvement [5]. Valid but incomplete data are included into the database but are not used in reports. This approach guaranties highly robust data for research but may considerably decrease the number of subjects and skew analysis due to missing datasets. Other registries persue a strategy similar to the ECFSPR, with inclusion of all submitted subjects but a standardized framework of measured to constantly enhance data quality, including audits, to strengthen clinical relevance of registry data [15–17].

An important finding of the Swiss data quality audit was the high variability between the centers concerning the availability of written ICs. In the 2022 re-audit, the number of missing ICs had improved from 15.5 % to 6.6 % after training of the CF centers; however, there was a persistently high variability across centers. In the ECFSPR, the availability of a valid IC must be confirmed in the data collection software ECFSTracker (tick box) by the local data manager during data collection.

In this data quality audit, we observed that some centers were not aware they should obtain written IC from all pwCF or their caregivers before registry data collection, despite this is a legal requirement. At the same time, providing individual patient data to the ECFSPR is compulsory for CFTR modulator therapy prescription in Switzerland, issued by the federal office of public health. The idea behind this prerequisite is that postmarketing data on outcome can be evaluated systematically. This may, in part, explain, why some individuals were included into the data base (legal requirement), but the ICs was not obtained beforehand for logistic or other reasons.

Shortcomings in collecting written informed consents from individuals before inclusion in a registry is a global and widespread issue, as demonstrated also by other groups (Boulton 2022). For clinical trials, ICs are collected thoroughly before a subject can be included, and their availability is strictly controlled during the study monitoring. Registries often do not have an established monitoring process and therefore the responsibility to collect and review ICs is in the hands of the local data managers only. This is bothersome especially when it comes to registry-based research and pharmacovigilance studies, where consent is mandatory. It has been shown that repeated audit cycles can largely improve compliance and consent rate to overcome these limitations [18].

As a limitation, the wide range of audited records per center between 25.3 % and 100 % could have skewed the results since comparison between a random sample and a total population per center was performed. However, the intention of this manuscript was not to compare the performance between the individual centers, but to get a nationwide picture of the data quality and the availability of important documents such as source data and ICs in a large national patient registry.

The findings of this Swiss registry audit on data quality have several implications for future registry data collection and might be of potential interest for other rare disease registries. Following the evaluation of the audits, several measures were implemented to improve data quality where it was lacking. General feedback on audit results were regularly released and discussed during the annual meetings of the Swiss CF centers, including training on data collection, data quality and completeness. CF centers were asked to obtain and provide missing source data on genetic information and missing ICs. CF centers received an individual written report on the results of the audit, together with information on how to improve their data collection in the following years. Repeated re-audits will be scheduled during the next years in all CF centers. Most importantly, national data managers were recruited and will perform all data collection in the Swiss CF centers in future, overcoming the problem of heterogenous qualification and training status as well as limited time resources of local data managers in the centers. These actions will enhance standardization, accuracy, consistency, and completeness of the annual registry data. A complete overview of all interventions to improve data quality in large, rare disease registries is depicted in Table 4. The determination of interventions and identification of their anticipated impact stated here were identified by the experience from the ECFSPR and Swiss national data quality audits, but also include suggestions from other reports on data quality in registries [3,4,7,8,12].

In conclusion, we demonstrate an overall good data quality in the Swiss CF registry with the few shortcomings mainly regarding the correct application of specific variable definitions. A varying number of missing informed consents were observed, which required direct feedback and re-audit. Specific feedback and teaching to centers, training of local staff, implementation of national data managers and repeated audits were implemented as a result of the onsite visits, to optimize data quality and provide robust data for scientific research, including registry-based pharmacovigilance studies.

Table 4
Interventions to improve and secure data quality in disease registries and their impact on data collection.

Intervention	Impact
Providing written SOPs and data coding documents / data dictionaries for data collection, variables, and variable definitions	Clear guidelines and references for data collection
Repetitive training of data managers on SOPs, variable definition and data collection	Securing of a sustainably high quality of data collection
Implementation of national data managers for a centralized data collection	Standardization of data collection
Regular data quality audits including individual feedback to centers and data managers	Re-evaluation and revision of performance, increase of consent rate
Data quality reporting (public and non-public)	Increase of awareness on data quality and areas for improvement
Precision and continual re-evaluation of unambiguous variable definitions	Reaction to observed haziness and uncertainties
Confirmation and transfer of source data, e.g., during transition from pediatric to adult center	Allocation of robust data
Built-in error recognition in the registry software	Quality control and feedback
Statistical error correction and data cleaning procedure after each data collection	Quality control and optimization

Funding

The project was funded by a data quality grant of Cystic Fibrosis Switzerland. The sponsor was not involved in study design, data collection and interpretation, writing of the manuscript and publication.

CRediT authorship contribution statement

Lara Wolf: Investigation, Data curation, Project administration, Writing – original draft, Writing – review & editing. **Jakob Usemann:** Formal analysis, Validation, Visualization, Writing – original draft. **Eugénie Collaud:** Investigation, Data curation, Writing – review & editing. **Marie-France Derkenne:** Investigation, Data curation, Writing – review & editing. **Reta Fischer:** Data curation, Resources, Writing – review & editing. **Maxime Hensen:** Data curation, Resources, Writing – review & editing. **Michael Hitzler:** . **Markus Hofer:** Data curation, Resources, Writing – review & editing. **Demet Inci:** . **Sarosh Irani:** Data curation, Resources, Writing – review & editing. **Kathleen Jahn:** Data curation, Resources, Writing – review & editing. **Angela Koutsokera:** Data curation, Resources, Writing – review & editing. **Rachel Kusche:** . **Thomas Kurowski:** . **Philipp Latzin:** Data curation, Resources, Writing – review & editing. **Dagmar Lin:** Data curation, Resources, Writing – review & editing. **Laurence Mioranza:** . **Alexander Moeller:** Data curation, Resources, Writing – review & editing. **Anne Mornand:** Data curation, Resources, Writing – review & editing. **Dominik Mueller-Suter:** Data curation, Resources, Writing – review & editing. **Christian Murer:** Data curation, Resources, Writing – review & editing. **Lutz Naehrlich:** Conceptualization, Investigation, Methodology, Validation, Writing – original draft. **Jérôme Plojoux:** . **Nicolas Regamey:** Data curation, Resources, Writing – review & editing. **Romy Rodriguez:** . **Isabelle Rochat:** Data curation, Resources, Writing – review & editing. **Alain Sauty:** Data curation, Resources, Writing – review & editing. **Macé Schuurmans:** Data curation, Resources, Writing – review & editing. **Michaela Semmler:** . **Daniel Trachsel:** Data curation, Resources, Writing – review & editing. **Anna-Lena Walter:** Data curation, Resources, Writing – review & editing. **Andreas Jung:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Kathleen Jahn, Andreas Jung, Philipp Latzin, Alexander Moeller, Christian Murer, Lutz Naehrlich, Alain Sauty and Jakob Usemann have received honoraria from Vertex Pharmaceuticals. Andreas Jung has received consulting fees and honoraria from EffRx Pharmaceuticals and Viatris. Philipp Latzin has received honoraria from CSL Vifor.

Acknowledgments

The authors would like to thank the ECFSPR for providing the templates to perform the Swiss national data quality audit, the Swiss cystic fibrosis centers for their collaboration and their engagement in the registry data collection and data quality, and all Swiss cystic fibrosis patients and their caregivers for allowing their data to be included into the registry. We also thank Fiona Beck for language editing.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jcf.2023.08.015.

References

[1] Barben J, et al. Zehn Jahre Neugeborenen-Screening auf zystische Fibrose in der Schweiz. Swiss Med Forum 2023;23(5):871–4. https://doi.org/10.4414/smf.2023.09321.

[2] Prang KH, Karanatsios B, Verbunt E, Wong HL, Yeung J, Kelaher M, Gibbs P. Clinical registries data quality attributes to support registry-based randomised controlled trials: a scoping review. Contemp Clin Trials 2022;119:106843. https://doi.org/10.1016/j.cct.2022.106843.

[3] Gliklich RE, Leavy MB. Assessing real-world data quality: the application of patient registry quality criteria to real-world data and real-world evidence. Ther Innov Regul Sci 2020;54(2):303–7. https://doi.org/10.1007/s43441-019-00058-6.

[4] Kodra Y, Weinbach J, Posada-de-la-Paz M, Coi A, Lemonnier SL, van Enckevort D, et al. Recommendations for improving the quality of rare disease registries. Int J Environ Res Public Health 2018;15(8):1644. https://doi.org/10.3390/ijerph15081644.

[5] Malenka DJ, Bhatt DL, Bradley SM, et al. The national cardiovascular data registry data quality program 2020. JACC state-of-the-art review. JACC 2022;79:1704–12. https://doi.org/10.1016/j.jacc.2022.02.034.

[6] Mikita JS, Mitchel J, Gatto NM, Laschinger J, Tcheng JE, Zeitler EP, et al. Determining the suitability of registries for embedding clinical trials in the United States: a project of the clinical trials transformation initiative. Ther Innov Regul Sci 2021;55(1):6–18. https://doi.org/10.1007/s43441-020-00185-5.

[7] Viviani L, Zolin A, Mehta A, Olesen HV. The European Cystic Fibrosis Society Patient Registry: valuable lessons learned on how to sustain a disease registry. Orphanet J Rare Dis 2014;9:81. https://doi.org/10.1186/1750-1172-9-81.

[8] Naehrlich L, Fox A, Krasnyk M, Orenti A, Zolin A, van Rens J, et al. The European Cystic Fibrosis Society (EFSPR) data validation program: accuracy and consistency of data. In: Poster presentation, ECFS annual congress; 2019.

[9] Zolin A, Orenti A, Jung A, van Rens J, et al. ECFSPR Annual Report 2021. 2023. https://www.ecfs.eu/projects/ecfs-patient-registry/annual-reports. last access 3.08.2023.

[10] https://www.ema.europa.eu/en/human-regulatory/post-authorisation/patient-registries; last access on 02.05.2023.

[11] https://www.ecfs.eu/projects/efcs-patient-registry/guidelines; last access on 02.05.2023.

[12] Pellen N, Guéganton L, Pougheon Bertrand D, Rault G. Lessons from the on-site quality audit of data transmitted to the French cystic fibrosis registry. Orphanet J Rare Dis 2018;13(Suppl 1):9. https://doi.org/10.1186/s13023-017-0750-x.

[13] Quanier, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J 2012;40:1324–43. https://doi.org/10.1183/09031936.00080312.

[14] Tingley K, Lamoureux M, Pugliese M, Geraghty MT, Kronick JB, Potter BK, et al. Canadian inherited metabolic diseases research network. Evaluation of the quality of clinical data collection for a pan-Canadian cohort of children affected by inherited metabolic diseases: lessons learned from the Canadian Inherited Metabolic Diseases Research Network. Orphanet J Rare Dis 2020;15(1):89. https://doi.org/10.1186/s13023-020-01358-z.

[15] Heinänen M, Brinck T, Lefering R, Handolin L, Söderlund T. How to validate data quality in a trauma registry? The Helsinki trauma registry internal audit. Scand J Surg 2021;110(2):199–207. https://doi.org/10.1177/1457496919883961.

[16] Manesen R, Mekler KA, Molobi TR, Tyiki AA, Madlavu MJ, Velen K, Charalambous S, van der Heijden YF. Data quality assessment of a South African electronic registry for drug-resistant TB, 2015–2016. PHA 2021;11(1):33–9. e-ISSN 2220-8372.

[17] Tan AC, Armstrong E, Close C, Harris IA. Data quality audit of a clinical quality registry: a generic framework and case study of the Australian and New Zealand Hip Fracture Registry. BMJ Open Qual 2019;8:e000490. https://doi.org/10.1136/bmjoq-2018-000490.

[18] Boulton C, Harrison C, Wilton T, Armstrong R, Young E, Pegg D, Wilkinson JM. Implementing large-scale data quality validation in a national arthroplasty registry to improve compliance: the National Joint Registry data quality audit programme. Bone Jt Open 2022;3(9):716–25. https://doi.org/10.1302/2633-1462.39.BJO-2022-0051.R1.