Cystic fibrosis patient registries: A valuable source for clinical research

Elliott C. Dasenbrook a, Gregory S. Sawicki b,*

a Cleveland Clinic Respiratory Institute, Cleveland, OH, United States
b Division of Respiratory Diseases, Boston Children’s Hospital, Boston, MA, United States

Received 19 December 2017; revised 26 February 2018; accepted 1 March 2018
Available online 16 March 2018

Abstract

Cystic Fibrosis (CF) patient registries are valuable data sources for researchers studying the natural history, treatment paradigms, and long-term health outcomes of individuals with CF. In this review, we discuss the role of CF patient registries in facilitating comparative effectiveness research, particularly evaluating therapies and variation in health care delivery. We also discuss the limitations of registry-based research, particularly indication bias, as well as statistical methods that can be used to address these issues.

Keywords: Cystic fibrosis; Patient registry

Contents

1. Introduction ........................................................... 433
2. Registry data as a valuable source for comparative effectiveness research ........................................................... 434
  2.1. Comparative effectiveness research: A framework ........................................................... 434
  2.2. Comparative effectiveness research to evaluate variation in healthcare delivery ........................................................... 434
  2.3. Comparative effectiveness research to evaluate therapies ........................................................... 435
  2.4. Comparative effectiveness research to evaluate genotype-specific CF treatments ........................................................... 435
3. Use of registries to facilitate prospective observational studies ........................................................... 435
4. Challenges with using registries for research ........................................................... 436
  4.1. Common limitations of registry-based research ........................................................... 436
  4.2. Addressing limitations of registry-based research ........................................................... 438
5. Future directions in registry-based CF research ........................................................... 439
Conflict of interest statement ........................................................... 439
Acknowledgements ........................................................... 439
References ........................................................... 439

1. Introduction

Broadly defined, a patient registry serves as an organized data collection tool incorporating clinical, sociodemographic, and other relevant health information obtained from multiple data sources across a health care system or systems [1]. Data included in registries should be collected using a standardized...
approach, and should be accessible to clinicians and researchers trying to understand and improve the health of a population. Applying a research lens, registry data is generally obtained through an observational study design, similar to classic cohort study designs in epidemiology. For decades, CF patient registries have been seen as a model for the development and use of patient registries of rare disease populations. Comprehensive registries for CF now exist throughout the United States, Canada, Europe, and Australia. The US Cystic Fibrosis (CF) Foundation Patient Registry (CFFPR), as an example, has become an important tool for health care providers, policy makers, and researchers for over five decades [2,3].

CF registries provide an avenue for the conduct of comparative effectiveness research (CER) and may also enable pragmatic, real-world clinical trials. In an accompanying review, Jackson and Goss discuss the origin of national CF registries, highlight the power of registries, explore issues related to international linkages of registries, and analyze how to sustain these registries into the future [4]. In this review, we highlight the promise and power of CF registries for such clinical research. In particular, we discuss the role of registries in conducting comparative effectiveness research of treatment approaches in CF and strengthening prospective observational studies. We also identify the challenges and limitations of registry-based research approaches, including statistical methods to address them.

2. Registry data as a valuable source for comparative effectiveness research

2.1. Comparative effectiveness research: A framework

Clinical research is traditionally designed using prospective designs developed with a single or limited set of questions and hypotheses in mind. In epidemiology, cohort studies are established to evaluate the impact of exposures on later health outcomes within a population. For rare outcomes, retrospective case-control studies are also used. When evaluating treatments or interventions, randomized controlled trials are considered the “gold standard”. It is apparent that for many research questions, particularly in small populations, such classic study designs are impractical, expensive, and for some subgroups may be impossible to execute. Comparative effectiveness research (CER), defined as the conduct of research comparing the benefits and harms of different interventions, management strategies, or treatments on health outcomes in “real world” settings, can help overcome some of these limitations [5,6]. Whereas clinical trials serve the purpose of establishing efficacy of treatments or interventions, the overarching purpose of CER is to provide an evidence-base for patients, clinicians, and policymakers about which interventions may be most effective under specific circumstances or with specific populations. CER often evaluates a broader group of health-related outcomes than clinical trials, often incorporates patient-reported outcomes and focuses on priority areas of interest to patient communities. In addition, since patients who enroll in clinical trials may differ from non-research participants [7], registry-based CER studies provide a methodology to evaluate outcomes across a larger set of patients. Registry-based CER research also enables the comparison of interventions and treatments in a “real-world” setting, and allows for evaluation of heterogeneity of effects across different subgroups; in CF populations, CER can include the study of patient populations that otherwise would not be eligible for clinical trials due to common trial exclusion criteria such as severely decreased lung function and presence of certain airway infections or co-morbidities for example.

2.2. Comparative effectiveness research to evaluate variation in healthcare delivery

Variation in treatment approaches between CF clinics has been identified for decades. CER addressing this variation in clinical care practices using registry data allows for evaluation of the impact of such variation on health outcomes. One of the earliest examples this research approach was a landmark study comparing outcomes in patient registries of those with CF receiving care in Toronto compared to Boston in the 1970’s [8]. During that era, patients in Toronto received differing nutritional recommendations emphasizing dietary fat intake, and they were found to have improved growth and survival despite similar lung function outcomes. In the modern-era, this study would be clearly considered CER. This study laid the groundwork for changes in nutritional CF care, and subsequent registry-based research has identified continued associations between better nutritional parameters and improvements in other health outcomes in several US and European registry analyses [9–11].

Variation in routine monitoring, frequency of hospitalizations, and use of chronic therapies has also been observed in CF registry-based studies. Although the Epidemiologic Study of CF (ESCF) was a North American based registry created primarily to evaluate long-term outcomes with dornase alfa therapy [12], data from ESCF also identified wide-ranging differences in practice patterns for routine monitoring visits and use of antibiotics and other chronic therapies [13,14]. Importantly, a 2003 analysis of ESCF showed significant differences in center-level outcomes based on these observed variations in care practices; namely centers that had increased numbers of monitoring visits and increased use of IV antibiotics had higher average lung function among their patients [15]. In a more recent UK CF registry study, a similar association was identified [16]. Waters and colleagues, using the Toronto CF Database, showed that subjects with pulmonary exacerbations treated with >14 days of antibiotics had a greater increase in lung function compared to those treated for ≤14 days, suggesting that peak lung function is not achieved in all patients within 14 days [17]. One important threat to interpreting possible center-level variation in care using registry data is a common lack of risk adjustment for the characteristics of the populations cared for at different centers. A recent UK registry based analysis showed that differences in median FEV1 across centers was minimal when adjustments for patient population characteristics were taken into account, concluding that apparent differences in outcomes were unlikely due to differing care practices [18]. Overall, CER approaches to critically evaluate specific components of “real world” CF care models and identify whether any particular
approaches are more effective in improving health outcomes within the CF population should still be developed.

2.3. Comparative effectiveness research to evaluate therapies

Registry-based CER can also be used to evaluate the real-world use and effectiveness of various chronic CF therapies. Over the past several decades, as more therapies have become available, the overall treatment complexity and treatment burden of CF care has increased [19,20]. Guidelines for routine CF care have provided recommendations for the use of many different chronic therapies [21], yet these guidelines do not address the clinical questions of whether recommended therapies are equally effective, additive, or could be interchanged. The Toronto CF Database was used to evaluate the impact of dosing and duration of inhaled antibiotics on eradication of initial Pseudomonas aeruginosa infection in children with CF in the real world setting [22]. Stanojevic and colleagues found similar rates of conversion to chronic infection between the treatment groups and both regimens were similar to eradication rates in controlled clinical trials. This study can impact decisions about eradication as the shorter course of inhaled antibiotic was associated with similar eradication rates as the longer course. Given a choice, most patients may opt for the shorter course. Chronic use of inhaled antibiotics is one area in which registry-based CER has been conducted. Looking at data from the 1996–2008 US CFFPR, consistent reported use of inhaled tobramycin among patients with CF and chronic Pseudomonas infection was associated with decreased mortality compared to no reported use [23]. This analysis supported guideline recommendations for chronic use of inhaled tobramycin. In 2010, after inhaled aztreonam was approved for the same indication, a subsequent US CFFPR analysis identified a large increase in the number of patients treated with more than one class of inhaled antibiotic [24]. This clinical practice change resulted in many more patients, of patients treated with more than one class of inhaled antibiotic aztreonam was approved for the same indication, a subsequent use of the comparator group could suffer from bias due to secular changes in clinical approaches or outcomes. Historical comparators can still provide valuable information, but researchers need to evaluate for such trends in order to justify selection of historical comparators. Finally, as discussed later on, some comparisons may suffer from confounding by indication, particularly in analyses of specific therapies.

2.4. Comparative effectiveness research to evaluate genotype-specific CF treatments

As the field of precision medicine in CF rapidly advances, another area where registry-based CER studies are important is in the understanding the impact of genotype-specific therapies of health outcomes of subpopulations of patients with CF. With the introduction of CFTR modulators for select groups of patients with CF in 2012, registry analyses should facilitate the evaluation of treatment responses to CFTR modulators, looking at longer-term safety and effectiveness in populations that were studied in clinical trials as well as those that may not have been included in such trials. Ivacaftor was first made available in the US in 2012 for patients age 6 years and older with CF carrying at least one copy of the G551D CFTR mutation. Analysis of the US CFFPR showed that initiation of ivacaftor therapy among eligible patients was quite rapid, with over 80% of eligible patients documented to be taking ivacaftor within a year after approval in the US [26]. Given the rapid uptake of utilization, analyses of longer-term outcomes of ivacaftor therapy require a different CF genotype subpopulation as a comparator group. Registry data can easily provide such data, and a 3-year analysis of lung function outcomes comparing data on G551D ivacaftor treated patients with matched delF508 comparators from the US CFFPR demonstrated that ivacaftor therapy slowed the rate of lung function decline [27]. Since 2012, indications for ivacaftor therapy have expanded to lower age ranges and to patients with an additional 37 CFTR mutations. Importantly, in the US, the most recent genotype-based ivacaftor approvals were based on in vitro and not clinical trial data [28]. Registry-based analyses in these additional subgroups will be even more critical in evaluating outcomes as ivacaftor use increases to other patient populations. Additionally, as new CFTR modulators are developed, studied in clinical trials, and potentially become available to patients, CF registries will need to collect data to facilitate CER analyses on longer-term outcomes. The first such analysis evaluating outcomes of lumacaftor-ivacaftor therapy, approved in the US as a CFTR modulator for patients with delF08 homozygous genotype, showed that compared to a matched US CFFPR cohort, patients treated with lumacaftor-ivacaftor also demonstrated a reduction in the rate of lung function decline [29].

An important consideration in any registry-based CER study on therapies is the selection and evaluation of appropriate comparator groups for the analysis. For example, the longer-term impact of ivacaftor on lung function decline was demonstrated using a delF508 homozygote group as a comparator to the G551D ivacaftor treated group [27]. Subsequent analyses showed that these 2 patient groups had similar rates of lung function decline prior to availability of ivacaftor therapy, supporting the use of the comparator group [30]. Additionally, a CER analysis based on a historical comparator group could suffer from bias due to secular changes in clinical approaches or outcomes. Historical comparators can still provide valuable information, but researchers need to evaluate for such trends in order to justify selection of historical comparators. Finally, as discussed later on, some comparisons may suffer from confounding by indication, particularly in analyses of specific therapies.

3. Use of registries to facilitate prospective observational studies

CF Registries can also be used to provide historical information in order to enrich the quality of prospective multi-center observational studies. A limitation to prospective observational studies is that important events in the year(s) prior to enrollment
are either not captured, or may be captured in an unreliable way. Failing to account for these unmeasured variables occurring in the years prior to the prospective study may result in biased results and conclusions. In CF, we are fortunate to have patient registries that reliably capture variables such as lung function measurements, hospitalizations, and microbiology. Linking the historical data of an individual enrolled in a prospective observational study significantly strengthens any associations detected during the study.

The standardized treatment of pulmonary exacerbations (STOP) study used this unique methodology of linking historical registry data to a prospective observational study [31–33]. A pulmonary exacerbation is one of the most common events in the life of an individual with CF and has a significant negative impact on survival, quality of life, and lung function [34]. Data from registries has supported the notion that there is wide variation in practices around pulmonary exacerbations among CF Care centers [35]. For example, the median length of intravenous antibiotic treatment for a pulmonary exacerbation in adults in the US ranges from 5 to 27 days [29]. Minimizing variation in practice has been shown to improve important patient outcomes [36].

In the STOP study, subjects that were enrolled into the study had 12 months of historical lung function measurements and use of antibiotics captured from the CFFPR. Through linkage of the CFFPR, the addition of the historical data allowed the investigators the important advantage of being able to understand the current pulmonary exacerbation in the context of “baseline” lung function and previous antibiotic use over the previous 12 months [31,32]. The historical data led to the following observations that could not have otherwise been gleaned from a traditional prospective study: 1) Approximately 20% of individuals with an exacerbation had an admission forced expiratory volume in one second (FEV₁) % predicted that was higher than their best FEV₁% predicted value recorded in the previous 6 months 2) At day 28 in the study, 35% of individuals did not recover 90% of their best FEV₁% predicted measurement in the prior 6 months, and 3) despite patients not reaching their previous lung function, most clinicians characterized treatment as a success. All 3 of these conclusions, which would not have not been able to be detected in a traditional prospective observational study that did not include historical data from a patient registry, will have a profound impact on future clinical trial study design of pulmonary exacerbations, especially endpoints determining success [33].

Therapeutic effectiveness in a real world setting is often evaluated in prospective observational studies. Ivacaftor had a tremendous impact in Phase 3 studies on lung function and pulmonary exacerbations [37]. Real world evaluations of therapeutic effectiveness are enhanced through linkage to a patient registry that can provide important data in the years prior to initiation of a therapeutic. Rowe and colleagues conducted a longitudinal multicenter cohort study of individuals with the G551D mutation to understand clinical status before and after initiation of ivacaftor [38]. The US CFFPR was used to provide historical microbiologic and hospitalization data. In the 12 months prior to ivacaftor initiation, approximately 25% of study participants were hospitalized and 55% had Pseudomonas aeruginosa detected in their sputum. In the 12 month period after starting ivacaftor, the percent of individuals hospitalized dropped to 10% and Pseudomonas aeruginosa was only detected in 35% of participants. The authors concluded that ivacaftor in the real world setting was associated with significant decreases in hospitalizations and Pseudomonas aeruginosa burden. These two associations could not have been detected without the addition of historical registry data to the prospective cohort study. Similar linked study designs are currently ongoing in patients homozygous for the F508del mutation, partial CFTR function, and/or absent CFTR function [39].

CF Registries have also been used in Phase 4 studies to investigate the microbiologic impact of an intervention. A 5-year prospective cohort study was conducted to understand the impact of aztreonam for inhalation on the susceptibility to aztreonam of Pseudomonas aeruginosa isolates [39,40]. Lower respiratory tract samples were collected from patients and the CFFPR provided the important clinical information. As more therapies become available for our patients, registries will play a larger role in understanding the impact of therapies, particularly long term safety and clinical effectiveness, after regulatory approval of a new drug.

4. Challenges with using registries for research

4.1. Common limitations of registry-based research

As discussed, registries are an important tool for CF research and have provided important information about therapeutic approaches and outcomes in real-world settings. However, as with any research methodology, there are important limitations to registry-based clinical research. First, not all registries are identical, and not all fields in registries are collected in a standardized manner. For example, the US CFFPR and Canadian registry use an encounter-based approach whereas the UK, and most European CF registries collect data on an annual basis. The timing of data collection can therefore result in limited data being available for certain research questions or outcomes, particularly if evaluating outcomes or treatment practices between different registries or systems. Efforts to harmonize and standardize CF registry data collection across countries should therefore be encouraged and are currently underway.

Registry data may also suffer from missing or incomplete information. A recent audit of the US registry encouragingly showed a low percentage of missing encounter data with a high degree of accuracy for key variables including demographics, lung function, and microbiology [2]. However, other fields, particularly those describing therapies or complications, may have a larger percentage of missing data. In general if missing data is <10% then there is unlikely to be significant bias [41]. It is critical for a researcher to assess the degree of missing data for any variables needed for an analysis. To address bias due to missing data, one analytic option is to restrict an analysis to records with complete data. However, such a method can introduce its own bias and is not desirable when the overall
N for any given analysis in a rare condition is already small. Since most missing data in CF registries is assumed to be missing at random, other statistical techniques to address missing data include performing sensitivity analyses using last observation carried forward or the multivariate imputation by chained equations method [42]. If the results of the sensitivity analysis are consistent with the main results, then the reader can be reassured that bias from missing data is most likely not significant.

Another particular challenge for registry-based comparative effectiveness analyses of therapies is the lack of information about adherence to any particular therapy or regimen. Although a registry may document the prescription or use of a particular therapy, there usually is no indication of whether an individual has actually obtained the medication or whether an individual’s adherence to the therapy is adequate. Indeed, poor adherence to chronic CF therapies, as defined by medication possession ratios, is well described [43]. Clinicians also may overestimate an individual patient’s adherence [44], and thus a registry may continue to document a prescribed treatment plan that is no longer being used. For these reasons, registry-based analyses on treatment effectiveness could result in biased estimates particularly if nonadherence rates are high. In addition, lack of clarity around regimen specificity can complicate CER analyses; this is certainly the case when evaluating daily respiratory therapy regimens or inhaled antibiotic regimens given the wide variation in clinical practice. These limitations generally stem from how registry data is collected, and thus may be overcome if a researcher understands the data collection process and tailors questions and analytic approaches to the available data. Incorporation of adherence data, either through linkages with administrative claims and pharmacy data or incorporation of electronically monitored adherence data could enhance patient registries to overcome this limitation.

Methodologically, a major challenge with CER using registry data is indication bias, also known as confounding by indication or confounding by severity [45]. The gold standard for evaluating the impact of a therapy is a randomized controlled trial. The power of randomization lies in the ability to balance unmeasured confounders between two treatment groups. Observational studies are not randomized, therefore, the indication for treatment may be related to the risk of a future health outcome (Fig. 1). In real-world clinical practice, patients who are sicker are more likely to receive more intensive treatment, so registry data capturing clinical practice is especially prone to this bias. This was illustrated in a retrospective US CFFPR analysis in 1999 which showed an increased risk of death among CF patients using inhaled tobramycin in 1998 even after controlling for other confounders [46]. More recently, a common scenario in CF clinical care would be an adult chronically infected with Pseudomonas aeruginosa who is taking an inhaled antibiotic every other month. (Fig. 2) A clinician or researcher might ask the question: what is the impact of adding a second inhaled antibiotic in the “off” month? One can therefore imagine a scenario with two patients. The first patient is doing well and symptoms and FEV1% predicted are at baseline (represented by the dark shapes). This patient is not likely to have a second inhaled antibiotic added to their regimen in the “off month”. In contrast, a second patient has had two pulmonary exacerbations in the last 6 months and has had a decline in their lung function (represented by the light shapes). This patient will be more likely to have a second inhaled antibiotic added to their treatment regimen. In follow-up, the patient on a second inhaled antibiotic will be more likely to have a lower FEV1% predicted. If a comparison between lung function outcomes between these two patients was performed, the results would suggest that a second inhaled antibiotic may result in worse lung function. This example illustrates how the imbalance in the severity of illness between the treatment groups introduces bias and may lead to

---

**Fig. 1.** Two sample patients from a patient registry where indication for treatment is related to the risk of a future health outcome. This is in contrast to a randomized controlled trial where treatment is randomly assigned.
inaccurate estimates of the relationship between a treatment and patient outcomes.

4.2. Addressing limitations of registry-based research

There are statistical techniques to reduce bias from the commonly encountered scenario of difference in severity of illness between study groups in observational research. There are several methods for addressing confounding by indication and the analytical field is rapidly evolving. Three such methods for controlling indication bias include restriction, propensity scores, and/or employing an instrumental variable [47–49]. Restriction is viewed less favorably due to the fact that it limits generalizability and in a rare disease like cystic fibrosis, other methods are preferred. Propensity methodologies and instrumental variables are becoming more commonly encountered in CF registry studies to evaluate the real world experience of CF care [27,50,51].

Propensity score matching can be utilized to increase the chances that patients across treatment groups are as comparable as possible and thus minimize confounding by indication. A propensity score was used to match CF patients with once copy of the G551D mutation and received ivacaftor to patients homozygous for the F508del mutation and lung function decline was estimated [30]. After creating two cohorts balanced on measured confounders, the authors demonstrated that lung function decline was slower by almost half in the G551D patients who received ivacaftor.

Returning to the example of inhaled antibiotics, propensity scores can be further explained (Fig. 2). As previously described, a commonly encountered situation in CF clinical practice is an individual with chronic respiratory tract *Pseudomonas aeruginosa* who is having frequent exacerbations despite treatment with an inhaled antibiotic every other month. The clinical question that arises is if the addition of a second inhaled antibiotic in the “off month” will provide additional benefit to a patient. A CF patient registry could be used to address this question by comparing patients in the registry that have stayed on monotherapy with an inhaled antibiotic to patients who had a second inhaled antibiotic added. This scenario is not a randomized controlled trial, so researchers need to address the differences between severity of illness between those on monotherapy and those who have had a second inhaled antibiotic added. All of the patients in the cohort will have varying measured clinical characteristics at baseline such as age, lung function measurements, microbiology, genotype, and/or medications. Propensity scores to predict the likelihood of being assigned to a treatment group (monotherapy (dark shapes) or combination therapy (light shapes) are generated for all subjects using the variables that are available from the patient registry. Each subject receives a propensity score. That score is represented in the figure by the size of the shape. Next, subjects from each treatment group are then matched on these scores to whittle the entire population down to a study cohort. As long as <10% of the population is excluded, the results are still considered valid. A significant strength of the various CF patient registries is that they are robust databases that allow the creation of high quality

Fig. 2. Sample CF Patient Registry that includes the population of patients on single inhaled antibiotic (dark shapes) and those on continuous alternating inhaled antibiotics (light shapes). The variety of the shape sizes reflects the variation in clinical characteristics between individual subjects in the registry. After the propensity scores are generated and patients are matched, the study cohort is created. Patients who cannot be matched (outside of the rectangle) are excluded.

propensity scores based on numerous well-characterized variables. Simulation studies have suggested that neither propensity or instrumental variable methodologies can fully overcome confounding by indication; however, a simulation study demonstrated that caliper matching on the propensity score led to one of the best estimates of treatment effect [52].

A third statistical technique to address confounding by indication is the use of instrumental variables. Instrumental variables have been used in one investigation to minimize bias from confounding by indication to understand the impact of inhaled tobramycin on lung function decline in CF [53]. A propensity score analysis was also performed in this study and only the instrumental variable analysis was able to successfully minimize confounding by indication. In the future, CF registry studies may consider using a center-specific variable as an instrumental variable (in the VanDyke and colleagues example, center-specific prescription rates were used) to help account for measured and unmeasured confounders and more accurately estimate the association between a treatment and an outcome [47].

In orphan diseases like CF, future assessments of existing therapies will need to leverage large prospective cohort registries to understand the real world impact. Comparative effectiveness research plays an important role as there are limited numbers of patients to conduct RCTs so some studies may not be feasible. Secondary data analyses do not prove causation, but the clinical impact of these real world studies may be the strongest evidence the CF community will have to evaluate the question of an intervention, thus efforts to minimize bias from confounding by indication is key.

5. Future directions in registry-based CF research

World-wide, CF registries have already had a tremendous impact on our understanding of the natural course of disease and on the identification of important risk factors impacting health outcomes. Registry-based research is also essential for evaluating long term effectiveness and safety of therapeutics. In the current era of big data, a key future direction for CF registries is the development and study of linkages across varied data sources. For example, there are differences in survival between countries and linkages between registries from countries will be vital to further understand these differences [54–56]. As these linkages become more common, especially internationally, protecting patients privacy when sharing registry data will be paramount [4]. In addition to comparisons across countries, CF patient registry research will benefit from including data from registries that contain complementary information. End-stage CF and the transition to transplant and outcomes post-transplant will be better understood through linkage with the scientific registry of transplant recipients [57]. CF patient registries do not contain information on whether our patients actually fill medications from a pharmacy, thus linking with claims databases will further augment the ability to assess adherence to chronic CF therapies. Linkages to other administrative or health system data sources could also allow for more rigorous evaluation of healthcare utilization, cost, and quality of care. With such linkages, CF registries will continue to be a powerful tool for research leading to improved health of individuals living with cystic fibrosis.

Conflict of interest statement

ECD and GSS have no relevant conflicts of interest to disclose related to this manuscript.

Acknowledgements

We would like to thank the individuals with cystic fibrosis and their families who have agreed to participate in registry research and everyone who has ever entered data into a cystic fibrosis patient registry.

References