UNIVERSITY OF LIVERPOOL

How many children recorded on the UK Cystic Fibrosis (CF) registry have an unclear diagnosis of CF following a positive newborn bloodspot screening (NBS) result and a potential designation of CRMS/CFSPID?

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Master of Philosophy by Ellie Russo

June 2021

Table of Contents

Abstract

Background

Infants with a positive newborn bloodspot screening (NBS) result that do not fulfil the criteria for a CF diagnosis are given the designation CF Screen Positive, Inconclusive Diagnosis (CRMS/CFSPID). We assessed the extent to which infants with an inconclusive diagnosis were added to the UK CF Registry, how many of the infants later converted a CF diagnosis and what data are available on follow-up sweat testing.

Methods

We considered all infants with a positive newborn screen test added to the Registry from 2008-2019. A CF diagnosis was confirmed by the presence of two CF causing *CFTR* variants or a sweat chloride value above 59 mmol/L. The harmonised definition of CRMS/CFSPID from the recent ECFS guidance (Barben *et al.* PMID 33257262) was used to classify individuals. The CFSPID designation was given to individuals who have a normal sweat chloride (<30mmol/L) and 2 *CFTR* variants (at least one with unclear phenotypic consequences) or an intermediate sweat chloride (30-59mmol/L) and one/no *CFTR* variant.

For infants with a probable CFSPID designation, conversion to a CF diagnosis was evaluated using repeat sweat test data, reclassified variants on the *CFTR*-2 database and the presence of clinical methods of diagnosis which may indicate a conversion due to the presence of symptoms. We also estimated the time between initial sweat test and first follow-up, and the number of follow-up sweat tests until age 5 and the number of follow-up tests over the age of 6 stratified by year of birth.

Results

In total, 2424 individuals were added to the registry with a positive newborn screening result between 2008-2019 and 1766 had at least one sweat test value recorded on the registry. 185 infants and children were identified that did not have a clear CF diagnosis, for whom a designation of CFSPID would be appropriate. There

was no reduction in numbers after 2014 (following Munck *et al*. PMID 25630966). Nine conversions were noted due to an increase in sweat chloride above 59mmol/L and less than five conversions were noted due to clinical symptoms.

49% of the individuals in the CFSPID population had only one sweat test performed and 35% had a repeat sweat test 31 days after diagnosis. There was no improvement in proportion of individuals receiving a sweat test at each age after 2014.

Conclusions

A significant number of infants and children are being included on the UK CF Registry without a clear diagnosis of CF. Numbers have increased slightly since the publication of the CFSPID designation, but this may be due to better recording of sweat test data. Average time to conversion is noted to be just over three years but sweat test data suggest that many individuals with CFSPID are not followed up according to international guidance which could delay the detection of conversions.

Acknowledgements

I would like to thank my supervisors Professor Kevin Southern, Dr Daniela Schlueter and Professor Ian Sinha for their unwavering support and guidance that I have received throughout this year. Professor Southern has guided me greatly through my project offering expertise and numerous opportunities to enrich my intercalated year and I have benefited greatly from his experience.

I would also like to thank Dr Schlueter for her patience and support which allowed me to significantly increase my statistical knowledge and to learn how to code from scratch. In a year that had the potential to be greatly affected by the COVID-19 pandemic, I felt thoroughly supported by my supervisors and every effort was taken to mitigate the effects. As a result, my confidence both conducting and presenting research has greatly increased, for which I am very grateful.

I am grateful to Dr Siobhan Carr, Dr Jane Chudleigh, Prof Iolo Doull, Pru Holder and Dr Maya Desai for their feedback and suggestions for how to improve this project. It was an honour to have so many clinicians and researchers interested in this project and their support has been immense.

The UK CF registry team, especially Rebecca Cosgriff and Elaine Gunn have offered invaluable assistance helping me navigate a large data set and their advice has shaped my work considerably.

Lastly, to my family, friends and loved ones, thank you for your unwavering support and patience; your constant support is greatly appreciated. Thank you!

List of Tables

List of Figures

List of Abbreviations

ASC	Adult stem cells
CBAVD	Congenital bilateral absence of vas deferens
CF	Cystic fibrosis
CFF	Cystic fibrosis foundation
CFFPR	Cystic fibrosis foundation patient registry
CFSPID	Cystic fibrosis, screen positive inconclusive diagnosis
CFTR	Cystic fibrosis transmembrane regulator
CFTR-RD	Cystic fibrosis transmembrane regulator related disorder
CHT	Congenital hypothyroidism
CRMS	CFTR-related metabolic syndrome
CUAVD	Congenital unilateral absence of vas deferens
DNA	Deoxyribonucleic acid
ECFS	European cystic fibrosis society
ECFS NSWG	European cystic fibrosis society neonatal screening working group
EGS	Extended gene sequencing
FE-1	Faecal elastase 1
FEV1	Forced expiratory volume in one second
ICM	Intestinal current measurement
IRT	Immunoreactive trypsinogen
MCADD	Medium chain acyl-coA dehydrogenase deficiency
NBS	Newborn screening
NHS	National health service
NPD	Nasal potential difference
PAP	Pancreatitis associated protein
PERT	Pancreatic enzyme replacement therapy
PKU	Phenylketonuria
PPV	Positive predictive value
SCD	Sickle cell disease
UK NSC	United Kingdom newborn screening committee
VUS	Variant of unknown significance
VVCC	Variance of varying clinical consequence

Aim of this work

- To explore newborn screening and the criteria for an infant to be given the designation of CRMS/CFSPID
- To determine the number of infants with an unclear diagnosis on the UK CF Registry and consider trends in the number of individuals with a potential designation of CFSPID before and after the introduction of the CFSPID designation in 2014.
- To detect conversions from CFSPID to CF due to increased sweat chloride, reclassified mutations or the presence of clinical symptoms.
- To examine sweat test data and determine if the CFSPID population are adequately followed up according to the recently published ECFS NSWG guidance.
- To inform national practice and the new CFSPID sub-Registry

Variant nomenclature

Throughout this thesis, the term “variant” will be used in place of “mutation” as it is no longer an acceptable term. A variant will first be described using the Human Genome Variation Society (HGVS) nomenclature with the legacy name in brackets. An example being c.1521_1523delCTT (F508del); after this, the legacy name will be used.

1. Introduction

1.1 Cystic fibrosis screen positive, inconclusive diagnosis (CFSPID) background

Since October 2007, the UK National Screening Committee (UK NSC) has recommended that all newborns are screened at five days old for cystic fibrosis using a dried blood spot sample (1). The blood spot is tested for levels of immunoreactive trypsinogen and if this is elevated two-stage genetic analysis is performed to detect *CFTR* variants. The final step for confirming a diagnosis is a sweat test to determine chloride channel dysfunction.

There is the possibility that a child may have a positive newborn heel-prick test due to a raised IRT, however, they do not fulfil the criteria to conclude a diagnosis of CF due to inconclusive *CFTR* genetic analysis and physiological testing (2, 3). These newborns have an inconclusive diagnosis which is known as CF screen positive inconclusive diagnosis (CFSPID) in Europe and *CFTR*-related metabolic syndrome (CRMS) in the US (4). CFSPID/CRMS was first introduced as a designation in 2014 (5).

The definitions were harmonised in a recent ECFS paper to create the CRMS/CFSPID designation. A child is given a designation of CFSPID if they fit the criteria for one of two groups (6):

- Normal sweat chloride (<30mmol/L) and 2 *CFTR* variants (at least one with unclear phenotypic consequences)
- Intermediate sweat chloride (30-59mmol/L) and one/no *CFTR* variant

Newborn screening is a valuable public health strategy, however it does result in the detection of children with CRMS/CFSPID which are seen as an unwanted outcome of screening. Although the harmonised definition uses the term CRMS/CFSPID, for the rest of this thesis the term CFSPID will be used to reflect the European population.

1.1.1 UK Cystic Fibrosis Registry

National registries help clinicians to understand their CF population, how effective their clinical practices are and to follow disease progression. Several countries have a national cystic fibrosis registry (7).

The UK CF registry was first established in 1995 and is sponsored and managed by the Cystic Fibrosis Trust. It has UK national health service (NHS) research ethics approval (8). The Registry collects data from annual review visits on people with CF in England, Scotland, Wales, and Northern Ireland and is one of the largest national databases (8, 9). It produces annual reports and registry studies which help characterise the CF population and provide evidence for commissioning care (8). It also allows longitudinal epidemiological studies of health outcomes in the UK CF population as well as international comparison studies.

There are 33 specialist paediatric CF centres and 28 adult centres in the UK which routinely collect data which results in high data acquisition as the registry is now estimated to cover 99% of the UK CF population (8, 10). Having a high proportion of the population present on the registry is important to allow for conclusions to be representative of the entire population and to reduce ascertainment bias (7).

Children with CFSPID should not be present on this registry as they do not have a confirmed diagnosis of CF. It is hypothesised that there are individuals with CFSPID present on the UK CF Registry especially as Ren et al noted that infants with CFSPID made up 15.7% of the US cystic fibrosis foundation patient registry (CFFPR) with a ratio of CF to CFSPID of 5:1 (11). In addition, there are inconsistencies with the identification of CFSPID with studies finding that 40.8% of individuals that fit the criteria for CFSPID were misclassified as having CF (3, 11). This could highlight discrepancies between the guidelines and clinical practice, which could result in unnecessary medicalisation due to the incorrect label of CF being applied to those who may never have been picked up clinically (12).

CF centres receive funding for individuals that they add to the CF registry. Clinicians may add individuals to the registry with an unclear diagnosis of CF as they are unsure about where to capture these individuals and want these children to be followed up and not 'lost' to the CF team in case of a conversion.

Therefore, this study is required to firstly identify those on the UK CF Registry that potentially have a designation of CFSPID and to consider the extent that they are added to the registry. The results of this study will inform the creation of a new CFSPID sub-registry that will help to better record this subset of patients.

The children added to the CF registry with CFSPID may receive full CF treatment, which can lead to unnecessary over-medicalisation and the risk of iatrogenic harm. Therefore, it is important that children with CFSPID are identified on the registry so that they can be managed appropriately as by definition they are well children.

Some clinicians may treat children with CFSPID prophylactically with standard CF care due to the risk of converting to a diagnosis of CF. There is anxiety surrounding conversion and apart from an intermediate sweat chloride it is not possible to predict those who will convert to a full diagnosis of CF.

However, this leads to difficult decisions about the division of resources which has become especially important with the development of *CFTR* modulator drugs. These are expensive medications, so it is important that children are only administered these treatments if they have a confirmed diagnosis of CF and a clinical presentation that will potentially benefit from these interventions.

As new ECFS guidelines on suggested follow up for these individuals have recently been published, it is important to understand how CFSPID patients are currently followed up by clinicians and whether this is in line with guidance (13).

In future, it is hoped that dedicated pathways will be created to optimise the care of individuals with CFSPID and to allow for individualised care based on their sweat

chloride, genotype, and other clinical markers. The results of this study will provide background for further development of care and help to inform clinicians of their progress with this condition and methods of improvement.

1.2 Cystic fibrosis background

1.2.1 Epidemiology

Cystic fibrosis (CF) is the most common inherited life-limiting disease of populations of northern European descent (14). It primarily affects those of European ethnicity (15) however CF is reported in all races (16). In the UK, approximately one in 2500 babies are born with CF and 10,655 people were present on the UK CF registry in 2019 (10).

CFSPID prevalence depends on the newborn screening algorithm used and will vary between programmes, however, frequency is increased in countries that use more extensive DNA analysis (3, 17, 18). If a protocol does not use DNA analysis and uses biochemical tests, fewer infants will have a designation of CFSPID. However this is at the expense of positive predictive value and sensitivity (13). CF to CFSPID ratio varies from 1.2:1 in Poland to 32:1 in Ireland reflecting the difference between programmes (19).

1.2.2 Molecular Genetics

CF is a monogenetic disease (20) caused by a variant in the cystic fibrosis transmembrane regulator (*CFTR*) gene on chromosome 7q31.2, which encodes a chloride ion channel protein (2). The *CFTR* transmembrane protein has five domains: two membrane-spanning segments that form the chloride channel pore, two ATP binding domains and a regulatory domain (16).

There are more than 2000 variants of the *CFTR* gene and the level of *CFTR* dysfunction correlates with the severity of symptoms (2).

The most common *CFTR* variant is c.1521_1523delCTT (delta F508) which is responsible for 50-80% cases (21) and is caused by a codon deletion of phenylalanine at position 508 of the *CFTR* gene. This results in a misfolded CFTR protein, which is degraded before it reaches the epithelial membrane (22). The other four most common variants are caused by a single base pair mutation (23).

CFTR variants are grouped into five (or sometimes six) categories based on the effect on the protein and *CFTR* dysfunction (24).

- Class I: truncated and non-functional protein
- Class II: misfolded protein
- Class III: full-length protein with gating defect
- Class IV: full-length protein with reduced conductance
- Class V: reduced amount of normal protein
- Class VI: reduced stability of protein

It is possible to produce a rough prognostication based on the class of variant with class I-III more likely to cause pancreatic insufficiency and severe lung disease, which deteriorates more rapidly (16, 20). Class IV-V variants are associated with a lower risk of pancreatic insufficiency, better lung function and survival (24, 25). However, there is a considerable overlap between the categories so it is not possible to accurately predict individual prognosis based on the class (20).

In the human lung, *CFTR* dysfunction results in less water in the periciliary fluid causing increased mucus viscosity (20). This is compounded by the knock-on effect on the ENAC sodium channel that reabsorbs sodium from the periciliary fluid reducing water content further (20). Mucus hyperviscosity causes a build-up in multiple organs particularly the respiratory and digestive tract (10). The thick mucus impedes cilial movement in the respiratory tract and hinders mucociliary clearance which predisposes to respiratory infection (20). Chronic infection with bacteria such as *Pseudomonas aeruginosa* and *Staphylococcus aureus* causes chronic sputum production, bronchiectasis, and lung destruction (16).

1.2.3 Classification of CFTR Variants

The *CFTR*-2 database has collected information from over 89,000 CF patients and provides information such as the classification of a *CFTR* variant, the likelihood of pancreatic insufficiency, lung function and pseudomonas infection rate for over 400 *CFTR* variants (26). The database categorises variants into four groups: CF causing, non-CF causing, variants of varying clinical consequence (VVCC) and variants of unknown significance (VUS) (11, 13). As of August 2020, there are 360 CF causing variants, 23 non-CF causing variants, 48 VVCC's and 11 VUS's listed on the *CFTR*-2 database (26).

Another database is CFTR-France that contains 16,819 variant records from 4,615 individuals which detail 736 different variants (27). These individuals have not only CF but also conditions associated with CFTR variants. These include CFTR-related disorder (CFTR-RD), fetuses with bowel abnormalities on ultrasound and asymptomatic compound heterozygotes (27).

Considering the classification of variants on the CFTR-2 database, a variant of varying clinical consequence can cause CF, a CFTR related disorder (CFTR-RD) or no disease. There are limited clinical data relative to variants of unknown significance and their classification is still unknown (28). Non-CF causing variants have been reported in the literature to cause CFTR-RD (28).

Variants are classified through a multistage process considering clinical implications, functional properties of the variant and epidemiology (28). The main criteria used is sweat chloride concentration as it is the gold standard test for diagnosing CF. A CF-causing variant is expected to cause an average sweat chloride greater than 60 mmol/L alongside another CF-causing variant (28).

The CFTR-2 database improves CF care by assisting with diagnostic, neonatal and carrier testing, and genotype correlation and indicating those who are eligible for new therapies (28). In addition, CF Source is a resource used by families and clinicians to provide information on the causes of CF and modulator therapy (29).

1.2.3 Clinical Presentation

Cystic fibrosis is a heterogeneous disorder (20) which presents with significant variation in clinical phenotype (1). It can involve several organs leading to clinical manifestations such as including lung disease with bronchiectasis, exocrine pancreatic insufficiency leading to malabsorption, hepatobiliary complications, and male infertility (2, 7).

Phenotype varies widely in the cystic fibrosis population and even individuals with identical genotypes may have diverse clinical presentations (7). Therefore, it is theorised that other factors such as the environment and epigenetics have a role in determining the severity of an individual's phenotype (7).

When considering the clinical trajectory of CF patients, the earliest symptoms detected are gastrointestinal symptoms such as, malabsorption causing faltering growth and failure to thrive (7). This is followed by respiratory symptoms during childhood and then later in life complications of CF will occur for example CF-related diabetes (7). Good care and new therapies can delay the occurrence of CF related complications and disease progression (7).

1.2.4 Therapeutics

Therapies for cystic fibrosis have rapidly improved and evidence-based medicine led by multidisciplinary teams has resulted in increased survival and life expectancy of patients with CF (7). There are three pillars for the management of cystic fibrosis: excellent nutrition, airway clear of infection and a healthy and active life (30).

Several aspects of CF management are personalised to the individual including airway clearance techniques and antimicrobial agents based on specific pathogens grown on culture (7). Alongside airway clearance techniques, adjuncts such as

hypertonic saline and azithromycin, have been shown to reduce the number of respiratory exacerbations (7).

Early intervention is effective in the lung to prevent more advanced lung disease such as bronchiectasis, however this approach is not as effective in other organs (7). Preventative medicine is not currently available for pancreatic insufficiency and replacement enzymes are required (7). However, studies have shown that it is possible that modulator therapy has a protective effect on pancreatic exocrine function (7).

Diet is an important part of CF management with a high calorie, high fat diet recommended with pancreatic enzyme replacement taken with every meal. In addition, fat soluble vitamin supplements are also recommended due to malabsorption (7).

CFTR variants are classified to determine whether they are eligible for treatment with a range of *CFTR* modulators which depends on the nature of the defect. For example, Kalydeco™ (ivacaftor) can be used in patients from four months old with a class III gating variant, Orkambi™ (lumacaftor/ivacaftor) can be used in a patient homozygous for F508 from 2 years old and Kaftrio™ (elexacaftor/tezacaftor/ivacaftor) is licensed for patients with at least one F508 variant from 12 years old (9, 24).

The development of new triple therapy, Kaftrio™ (ivacaftor, tezacaftor, elexacaftor) has the potential to considerably improve outcomes for a significant proportion of patients with CF.

1.2.5 Newborn Screening Background

Universal cystic fibrosis screening was recommended by the Minster of Public Health England in April 2001. However, newborn screening for CF was already occurring regionally in Wales, Northern Ireland, and certain parts of England.

Scotland introduced cystic fibrosis newborn screening in February 2003 and subsequently a universal UK programme commenced in October 2007.

Thus, all newborns are screened at five days old using a heel-prick blood test to detect phenylketonuria (PKU), congenital hypothyroidism (CHT), sickle cell disease (SCD), medium-chain acyl-CoA dehydrogenase deficiency (MCADD) and cystic fibrosis (CF) (1).

A population-based screening programme for cystic fibrosis first became feasible when Crossley identified elevated levels of immunoreactive trypsinogen (IRT) in bloodspot samples of CF positive neonates (31). The *CFTR* gene was identified in 1989 and DNA analysis is used in cystic fibrosis NBS alongside IRT (32, 33).

The screening programme follows Wilson and Jungner guidelines which are considered the gold standard of screening assessment (17, 32). This is centred on the concept that a condition should be identified at an early stage when the patient is asymptomatic and furthermore, there is a treatment available that will improve the outcomes of the patient (32). Screening is advantageous as treatment can be initiated early during the short pre-symptomatic phase to avoid some of the complications that occur if cystic fibrosis is left untreated in the neonatal period. These complications include malnutrition and electrolyte imbalances such as, hyponatraemic dehydration and detection and early treatment of respiratory infections (2, 34).

In addition, screening provides an opportunity to recruit children into trials before the onset of lung disease and has the possibility of benefitting both the patient and the wider CF community (34). An early diagnosis of CF facilitates the initiation of *CFTR* modulator therapy, which can correct the underlying defect (35, 36). These therapies may prevent a child from developing progressive exocrine pancreatic damage and insufficiency and can improve growth (9, 17, 25). CFTR modulators can reverse pancreatic insufficiency when started early in life (36).

Around 70% of babies born with CF are now diagnosed by newborn screening and they are often asymptomatic when diagnosed reducing differential time to diagnosis (10, 33). Newborn screening is also cost-effective as screening costs are offset by savings made due to fewer interventions being necessary (37).

In contrast, newborn screening was not shown to reduce health inequalities when considering socioeconomic class. It was hypothesised that reducing time to diagnosis would help to reduce inequalities, however this result does show that there are a lot of factors beyond early diagnosis and clinical care that influence disease development (10).

An ideal screening programme has high sensitivity so those with disease are detected, whereas high specificity is required to discount those who will not develop disease (33). It is also important to design the programme to have a high positive predictive value (PPV) to yield low numbers of false positives, which can have a psychological impact on parents and also low numbers of false negatives which can lead to a delayed diagnosis and late initiation of treatment (12).

In 2016, there was a newborn screening programme in 21 European countries, however, there is considerable variability in the protocols used to diagnose CF (38). The differences between the protocols used are due to several reasons. These include different healthcare systems, the absence of a universal tool to evaluate performance, cost of different programmes and barriers to change in countries with established programmes (38). However, harmonisation of newborn screening programmes is not desirable as there are significant differences between populations (39). In the UK, there is a consistent approach of IRT level followed by genetic analysis and a confirmatory sweat test. The unique characteristics of the UK NBS protocol are outlined in more detail later (10).

1.3 Newborn screening protocol

1.3.1 Antenatal testing

In-utero diagnosis of CF can be made through antenatal carrier screening (20). First, the carrier status of the couple needs to be determined and this can be done in two tiers where one partner is screened first for common variants; then if a variant is detected, a larger panel can be screened for in the other partner to make sure rarer variants are not missed (16). If both parents are known heterozygotes it may be useful to determine whether their fetus is affected by CF (20).

Screening during pregnancy allows for elective termination of pregnancy if the fetus is found to be affected (16), however, with the development of modulator therapy continuation of the pregnancy may often be preferred (7, 40). An early pre-natal or post-natal diagnosis could also prove favourable if modulator therapy is approved for use in infants (7).

1.3.2 Immunoreactive Trypsinogen

The first step in the diagnosis of CF via newborn screening is the analysis of bloodspot samples for levels of IRT, which is produced by the pancreas. This is a shared first stage across all cystic fibrosis screening programmes as IRT is raised for the first few weeks in babies with CF (12, 38).

An elevated IRT level can indicate damage to the exocrine function of the pancreas as obstruction of pancreatic ducts leading to enzymes being drained into the intestine (7, 17). IRT level has been shown to correlate with the severity of the *CFTR* function and phenotype (14, 41). Even milder *CFTR* dysfunction affects the pancreas as CF carriers and patients with pancreatic sufficient CF phenotypes also have a raised IRT (17). However, Ooi found higher IRT levels and a more rapid decline associated with more severe *CFTR* variants (class I-III) (14). The rate of decline of an IRT level is variable and IRT is no longer reliable as a cystic fibrosis diagnosis marker at 8 weeks of age (1, 42).

There are two main ways to evaluate an IRT result; they are using a fixed cut-off of a certain range of IRT values or floating cut-off to send a set percentage of high-risk

samples for further testing (38). A fixed cut-off may include IRT values of 60-90 ng/ml whereas a floating cut off may range from 95.0th to 99.5th centile (38). The screening programme in the UK uses IRT-DNA-IRT protocol (43) and if the first sample meets either cut-off, then two-stage variant analysis is undertaken (6).

There is an inherent trade-off when deciding the cut-off values for IRT, as too high a cut off results in missed cases which are false negatives. However setting the value too low, although increasing sensitivity, results in a significant number of unaffected carriers and inconclusive diagnoses being detected (1, 44). As the specificity of only a single IRT measurement is low, it is used as part of a two or three-stage programme along with DNA analysis and a second IRT measurement (12). There is variation in second tier tests between programmes (5). However most screening programmes use *CFTR* gene analysis as the second stage after initial IRT testing (33).

A second IRT level might be used if no variant is initially detected during *CFTR* gene analysis, but the first IRT was considered very high, greater than 99.9th centile (17). This avoids missing cases with rarer *CFTR* variants that may not be detected by the variant panel and is often referred to as a "safety net" strategy (1, 17, 38). If the IRT levels are still high, then a preliminary diagnosis of CF is made (1). This strategy reduces false negatives which increases sensitivity and a second IRT sample has a stronger positive predictive value as serum IRT declines more rapidly in false positive cases (40, 43).

CFTR dysfunction is not the sole reason for the increase in IRT in a newborn and this can result in false positive screening results. Other causes of a raised IRT include prematurity (especially <29 weeks' gestation), congenital infections, renal failure, bowel atresias, nephrogenic diabetes insipidus and chromosomal abnormalities (14, 17, 43). IRT levels are also elevated in black neonates compared with white, causing a greater proportion of positive NBS tests (11).

In contrast, the neonatal emergency meconium ileus is often associated with a low IRT value and can therefore lead to false negative results (38). Meconium ileus is a type of bowel obstruction that occurs when a newborn stool is stickier than normal due to abnormal pancreatic enzyme secretion and is not passed (16, 45). It is a highly suggestive feature of cystic fibrosis affecting 20% of neonates with the condition and is associated with more severe *CFTR* variants (20). It can be occasionally seen on antenatal ultrasound between 17th and 22nd weeks of gestation as increased echogenicity of the fetal intestine however, this is not a sensitive or a specific sign as there are numerous causes other than CF (7, 20). Meconium ileus presents early therefore patients are diagnosed clinically and not usually through newborn screening programmes (38).

1.3.3 *CFTR* Variant Analysis

Globally, the initial number of variants screened for ranges from four to 644 and there is huge variability between programmes due to differing prevalence of *CFTR* alleles in different populations (38). For example, F508del prevalence in the general population is 70% in Germany but 25% in Turkey (9).

Certain countries like the UK use two-stage variant analysis as this minimises carrier detection and those with milder variants which may not be clinically significant and therefore do not require a sweat test (1). However, as first stage testing only screens for the most common variants, rare mutations may not be picked up by newborn screening and therefore, a delayed diagnosed may be made clinically (43).

Extended genotype analysis, using a large panel of variants, improves the specificity of newborn screening compared to an IRT only programme (13, 18). However, improving one aspect of screening performance can often harm another aspect. Although DNA analysis improves positive predictive value with fewer false positive infants undergoing sweat testing it does result in more carriers and inconclusive diagnoses, leading to clinical dilemmas on how to manage these patients and length of follow up (13, 43).

However, non-mutation specific analysis may be suitable for populations with considerable genetic variability. There are currently 1500 rare variants that have been discovered but not assessed due to lack of data which could result in an unknown phenotype even with the development of the *CFTR*-2 database (13, 17, 44).

Course stated that variants of unknown significance (VUS's) may just be a normal variation in *CFTR* genotype so it is argued by Wilfond that newborn screening should only identify variants that cause pancreatic insufficiency (33, 46). There is an opportunity to increase specificity if variant panels are restricted to only CF-causing variants (38); however most programmes still screen for VVCCs to avoid missing cases with differing phenotypes (13).

For example, c.350G>A (R117H) variant has variable clinical consequence showing low penetrance for clinical CF, causing symptoms of CF in some and single organ manifestation of *CFTR*-related disorder (*CFTR*-RD) in others (11, 41). The severity of phenotype depends on the length of poly T tract before exon 9, 5T variant is CF-causing whereas 7T or 9T variants are a VVCC, causing *CFTR*-RD or CFSPID (13). R117H is approved for treatment with ivacaftor modulator therapy, but children with CFSPID and a normal SCC should not receive this therapy (13).

There is a risk that carriers have a second abnormal allele not detected by the variant panel, especially considering different ethnic populations with different *CFTR* variants (1, 17). Therefore, further IRT testing may be required as a falling IRT level can be indicative of carrier status (38). Detection of carriers is generally considered to be an unwanted outcome of screening and a limited DNA panel reduces the detection of carriers. It impacts the extended family of a patient (47) due to cascade testing as relatives of reproductive age may undergo gene analysis to determine their status; however, it does help with identification of high-risk couples within families (17, 44).

1.3.4 Pancreatitis Associated Protein

Another biomarker that is used by four European countries alongside IRT is pancreatitis associated protein (PAP) (48). It is a secretory protein which is detected when there is pancreatic injury, however, a rise in PAP is not CF-specific and there is no evidence that shows a correlation between PAP level and severity of disease (48).

Adding in PAP testing after initial IRT and DNA analysis (IRT/PAP/DNA) can result in fewer carriers and fewer cases of milder forms of disease, such as CFSPID being detected (49). This improves the NBS protocol by increasing positive predictive value however this is at the expense of sensitivity (13, 50).

Using an IRT/PAP strategy can be useful in countries with high ethnic diversity and considerable heterogeneity of *CFTR* variants as PAP alongside a safety net IRT value can remove the diagnostic uncertainty that comes with a small panel of *CFTR* variants (49). In addition, it can reduce the number of sweat tests that need to be performed and leads to an earlier diagnosis than using a second IRT measurement at 21 days (48, 49).

Figure 1 highlights different national newborn screening protocols in European countries with national programmes (36). The countries without a colour often have regional programmes which can have very different protocols.

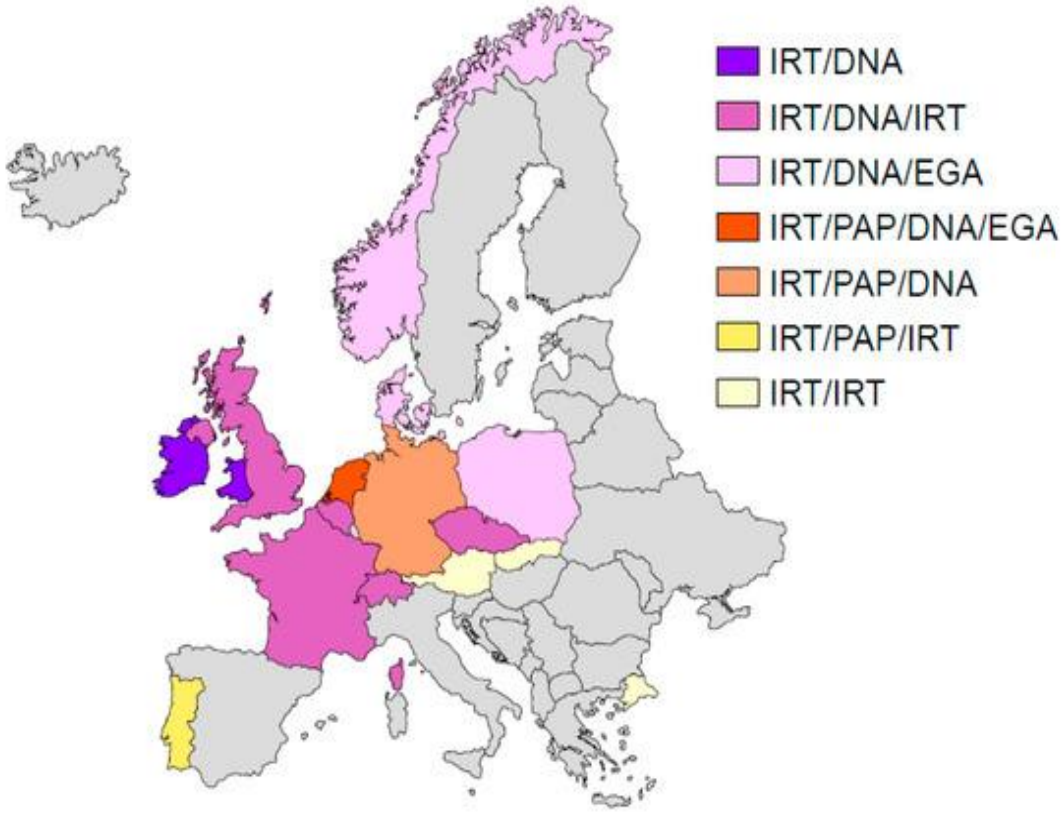


***Figure 1**: National newborn screening algorithms used in Europe (36)*

1.3.5 Characteristics of the UK Screening Programme

The UK (apart from Wales) uses an IRT-DNA-IRT protocol (51). To start, the first IRT has a very high cut-off with the 99.5th centile continuing onto DNA analysis. This leads to less individuals continuing onto DNA analysis. The bloodspot sample that is tested for IRT is taken on day 5 of life, which is later than other countries. This is done to reduce the risk of false positives as IRT can be raised in conditions other than cystic fibrosis. The combination of both of these measures leads to fewer individuals needing DNA analysis and discounts individuals unlikely to have CF early in the newborn screening protocol (51).

Genetic testing in the UK is done in two stages; first testing for the four most common *CFTR* variants in the UK population associated with severe disease. These four variants cause more than 80% of disease and are F508del, c.1652G>A (G551D), c.1624G>T (G542X) and c.489+1G>T (621+1GT) (1). If a child tests positive for one

variant, then a further larger panel of variants is tested. The results of this second panel determines whether a child has a second CF-causing variant leading to a CF diagnosis or only has the first variant meaning they are a probable CF carrier (1).

If one variant is present on either panel then the infant undergoes a second IRT measurement at 21 days. Carriers can have a high initial IRT and second IRT is often normal preventing carriers from requiring a sweat test.

If no variant is found but the individual has an ultra-high IRT then a preliminary diagnosis of CF is made and a sweat test is performed (1). This acts as a safety net to not miss individuals with rarer and possibly non-European *CFTR* variants. This strategy reduces false negatives which increases sensitivity and a second IRT sample has a stronger positive predictive value as serum IRT declines more rapidly in false positive cases (40, 43). For the UK protocol, if the second IRT is low, and one variant is detected then the individual is considered a carrier and does not undergo a sweat test (51).

However, Wales do not perform a second IRT measurement. They instead follow an IRT-DNA protocol and if one variant is detected, infants undergo sweat testing. This means that Wales' protocol has a lower positive predictive value (PPV) than the UK protocol due to many sweat tests being performed.

The UK CF newborn screening algorithm is shown in **Figure 2** and outlines the cut-offs used for IRT measurement and the different outcomes of the newborn screening protocol (1).

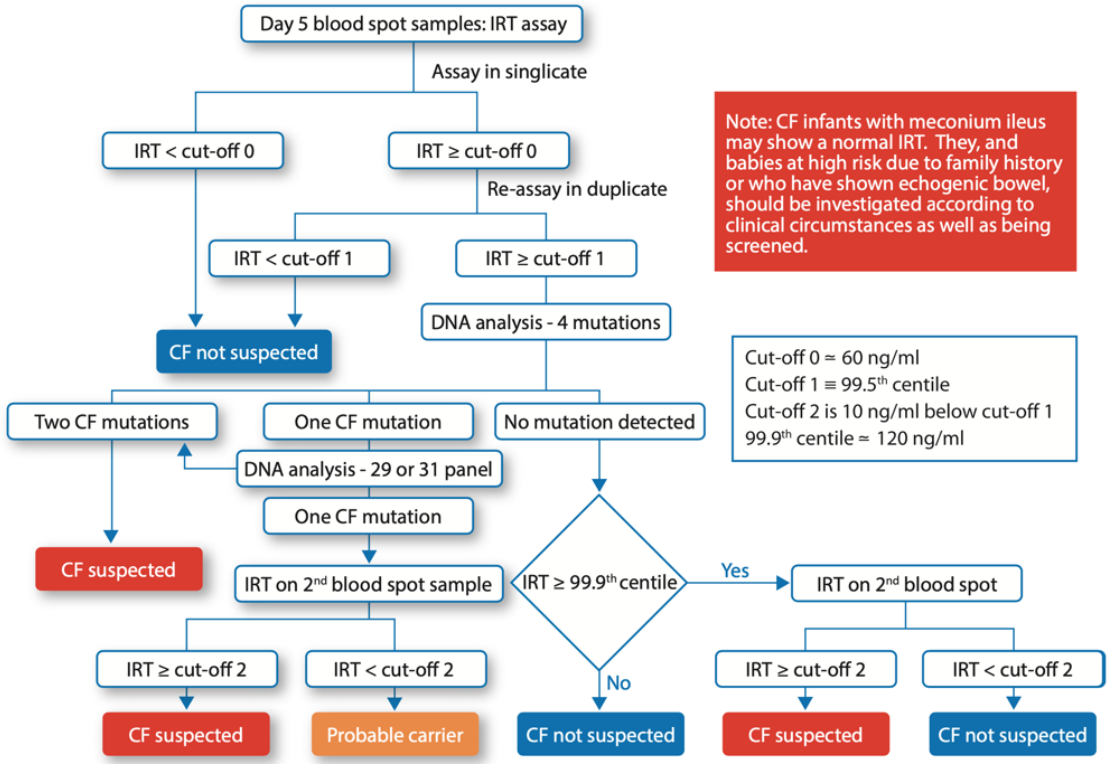


***Figure 2**: UK CF newborn screening algorithm (1)*

1.4 Diagnostic Evaluation

After a positive newborn screening result, clinical evaluation of the patient and a sweat test are important to confirm a diagnosis of CF.

1.4.1 Sweat Chloride Testing

To confirm a diagnosis of cystic fibrosis, the gold standard is a sweat test to demonstrate *CFTR* chloride channel dysfunction and *CFTR* variant phenotype as it is the most sensitive and specific diagnostic tool (7, 20, 33). A normal sweat chloride level is below 30mmol/L and is unlikely to be cystic fibrosis (2). An intermediate sweat chloride is 30-59mmol/L and often requires further DNA analysis and clinical evaluation (50). A positive sweat chloride is greater or equal to 60mmol/L, which is diagnostic for cystic fibrosis (41).

However, it is possible for a child with a normal sweat chloride test to have two CF-causing variants (43), for example, the variant R117H is associated with a normal or equivocal sweat test result at first which later become abnormal (1).

Farrell believes it is advisable to perform genetic analysis on those with an intermediate sweat chloride (2) however it is not uncommon for genetic analysis to result in variants of varying clinical consequences which can lead to uncertainties over diagnosis (20).

There are 3 main circumstances in which a presumptive diagnosis of CF is made (1, 2):

- Positive CF newborn screening showing 2 CF-causing *CFTR* variants
- Positive CF newborn screen and clinical signs and symptoms of CF
- Meconium ileus with or without positive newborn screen

Other circumstances include an antenatal diagnosis or an individual being a sibling of an existing patient.

1.4.2 Nasal Potential Difference

Electrophysiological tests can be used to assist in the diagnosis of CF in atypical cases where *CFTR* dysfunction is suspected. These include nasal potential difference (NPD) and intestinal current measurement (ICM). They are performed in very highly specialised CF centres and are not routinely used in clinical practice (19).

Nasal potential difference (NPD) measures transepithelial ion transport of sodium and chloride in the nasal airway which reflects *CFTR* function and phenotype (52). These nasal airway abnormalities can reflect *CFTR* abnormalities in lower airways allowing channel function to be evaluated. NPD is challenging in pre-school children and is not routinely undertaken to diagnose CF (13, 52, 53).

1.4.3 Intestinal Current Measurement

Intestinal current measurement (ICM) records ion transport in rectal biopsies after stimulation with chloride secretory agents and can be useful to confirm a diagnosis of CF if a sweat chloride test has resulted in an intermediate value (2, 53). ICM is

considered to be more feasible in children than NPD however neither are used routinely in clinical practice (7).

As *CFTR* genotyping has advanced substantially, the diagnostic role of electrophysiological tests such as NPD and ICM is now unclear (13). However, these tests may have a use in determining variant pathogenicity especially considering variants with equivocal clinical phenotype (40).

1.4.4 Organoids

Preliminary evidence has shown that intestinal organoids may have a role in CF care (13). Organoids are a three-dimensional collection of organ specific cells that are comparable to the cells in-vivo (54). The most widely used cell types for organoids are embryonic and induced pluripotent stem cells (IPSCs) as they have the potential to differentiate into almost any cell type; however, adult stem cells (ASCs) can also be used (54).

Organoids can be used for disease modelling, drug screening and personalised medicine and are theorised to be used in CF to predict clinical benefit from *CFTR* modulators (54). Overall, currently there is insufficient evidence to recommend the use of intestinal organoids routinely for diagnosis of infants with CF(13).

1.4.5 Faecal Elastase

Faecal pancreatic enzymes can also be used to determine the presence of exocrine pancreatic insufficiency and the most common enzyme measured is faecal elastase-1 (FE-1) (20). Pancreatic insufficiency is detectable at birth in 60-80% of newborns with CF and can cause symptoms of fat malabsorption such as malnutrition and failure to gain weight (20).

FE-1 can be used if a sweat test is inconclusive where one CF causing variant has been identified. Reduced FE-1 can be considered presumptive of a exocrine pancreatic insufficiency and a diagnosis of cystic fibrosis (13) . This avoids the need

for using more extensive DNA analysis which can confuse if the phenotype of a variant is not well characterised (13). Testing faecal elastase can also increase the timeliness of initiation of pancreatic enzyme replacement therapy (PERT) improving anthropometric markers such as weight and preventing growth stunting which is not seen in pancreatic sufficient infants (25, 39).

1.5 Cystic Fibrosis Screen Positive, Inconclusive Diagnosis (CFSPID)

CFSPID prevalence depends on the newborn screening algorithm used and is variable between programmes, however, frequency is increased in countries that use more extensive DNA analysis (3, 17, 18). If a protocol does not use DNA analysis and uses biochemical tests, fewer infants will be identified as having CFSPID, however this is at the expense of positive predictive value and sensitivity (13).

It is important to note that children with CFSPID do not have cystic fibrosis and their phenotype is varied so it is possible for patients to be asymptomatic at diagnosis and never display CF symptoms (5). A subset of CFSPID patients report clinical symptoms such as wheeze, cough, constipation, and abdominal pain but these are much less frequently reported than those with CF (41). Despite the possible lack of symptoms, children should be referred to a specialist CF centre for clinical evaluation and to establish a future management plan and follow-up schedule as the natural history of CFSPID is still unclear (2, 33).

1.5.1 Monitoring and management of CF screen positive, inconclusive diagnosis

New guidance on the monitoring of children with CFSPID has been published by the European CF Society Newborn Screening Working Group (ECFS NSWG) as previously there was little consensus on the optimal follow up and management of CFSPID and debate surrounding the length of follow up (13, 33). It is not yet clear whether screening for CFSPID is advantageous and has clinical benefit however it seems that most individuals with CFSPID will remain asymptomatic.

Following a designation of CFSPID, a child should undergo clinical assessment alongside *CFTR* analysis, sweat testing, and measurement of FE-1. The frequency that a child is seen in clinic should depend on their well-being and the perceptions of their parents and carers. Visits can decrease if a child is growing and gaining weight suitably and their family gaining in confidence (13).

Sweat tests should be repeated at six months and two years followed by yearly testing until age 6, to monitor sweat chloride levels. It is also necessary to check for further updates on the *CFTR*-2 database regarding the reclassification of *CFTR* variants that could lead to a change in diagnosis (13).

At six years, it is proposed that a more extensive clinical evaluation is undertaken which will include sweat testing, pulmonary function testing, chest imaging and FE-1 measurement. Further evaluation is done at this age as data from prospective studies has shown that if tests are normal, a child is unlikely to convert to a diagnosis of CF hereafter (55). There is then the possibility of discharge from the CF clinic with the responsibility of care being transferred to the primary care physician (55). A further assessment by the CF team is recommended in adolescence to discuss issues such as fertility and future reproductive decision making (13).

There is a careful balance between overmedicalizing or undertreating a patient with CFSPID. However, there are no studies that recommend the routine use of CF therapies for CFSPID patients due to the risk of iatrogenic harm (13). Ren et al found that infants with CFSPID were significantly less likely to be prescribed CF therapies such as pancreatic enzyme replacement therapy (PERT) and dornase alfa but are more likely to be prescribed bronchodilators (11). However, early proactive treatment has not been shown to improve long-term clinical outcomes for patients with CFSPID (3). Therefore, a less intensive approach to managing children with CFSPID is recommended with an emphasis on clinical monitoring rather than exposure to unnecessary investigations and interventions (3, 55).

Furthermore, if CFSPID patients are placed into CF outpatient follow-up clinics, this can put a strain on these services especially when children may not require this level of follow up and can also place a strain on families (33). It is also important to note that families with children who are clinically well may be lost to follow up as they disengage with clinical care and therefore data on CFSPID cohorts may be biased towards more symptomatic children (17).

Infants with CFSPID have been shown to have lower initial IRT values than infants with a CF diagnosis, which potentiates the use of IRT to predict the likelihood of converting to a diagnosis of CF (14). However, it remains unclear, at the individual level, whether IRT can predict which infants with CFSPID will convert to a CF diagnosis (13). Furthermore, sweat chloride levels have also been found to be lower in children with an inconclusive diagnosis with many in the normal (<30mmol/L) and intermediate range (30-59mmol/L) (41).

At birth, there are no significant differences in weight between CF and CFSPID cohorts but by the time of initial assessment, the CF cohort has a significantly lower weight and height (41). This is due to children with CF not producing sufficient pancreatic enzymes to assist with digestion leading to malabsorption (31). At diagnosis, most CF patients are pancreatic insufficient whereas CFSPID patients are pancreatic sufficient (18).

1.5.2 Conversion from CF screen positive, inconclusive diagnosis to cystic fibrosis

CFSPID cohorts are closely monitored by CF specialists due to the risk of developing CF-related conditions and converting to a diagnosis of CF (5, 6). Although there is a balance between overmedicalising and undertreating, sweat test follow up provides an opportunity to prevent illness occurring in later life which is associated with worse outcomes (7).

It is possible for a designation of CFSPID to convert to CF if a repeat sweat chloride test is raised, variants are reclassified as CF-causing, or a patient becomes

symptomatic (5). If a diagnosis is changed due to new knowledge of a CF-causing variant from *CFTR*-2 database, this may not represent a change in the clinical picture of the individual rather more knowledge being acquired on the phenotype of the variants (13).

It is currently estimated that conversion from CFSPID to CF is between 6% and 48% and there is considerable variability between studies and populations (13, 33). However, the study that noted the high conversion rate used definitions of CRMS/CFSPID that are slightly different to those included in the ECFS guidelines as there was limited DNA analysis (only testing for F508del) and only individuals with an intermediate sweat chloride were included (56).

The variability results in an inability to predict which individuals accurately and consistently will convert to CF (33). Discrepancies in conversion rate are thought to be due to different definitions of inconclusive diagnosis, different interpretations of *CFTR* variants and differing durations of follow up (13). When trying to predict what individuals will convert to CF, Ooi found that symptoms in the first 3 years of life prior to a diagnosis is not a significant discriminator (33).

Different methods of diagnosing CF also cause discrepancies in conversion rate with extended gene sequencing leading to 5% of infants with CFSPID converting to CF whereas the rate for conversion using biochemical methods is higher at 25-40%, however there are fewer individuals with CFSPID identified through biochemical methods.

A child who has a diagnosis conversion often presents with a milder, atypical form with a normal or intermediate sweat test (3). Sweat chloride should be monitored even in asymptomatic patients, as patients with variants of varying clinical consequence (VVCC) may show an increase in sweat chloride over time and lead to a conversion from CFSPID to CF (18, 33). Increased attention should be placed on

those CFSPID patients with intermediate sweat chloride levels (30-59mmol/L) as they are more likely to convert to a diagnosis of CF (19).

However, even when considering the same *CFTR* variant, the outcome can be variable; for example, the VVCC c.3454G>C (D1152H) presents in some with pancreatitis and isolated bronchiectasis later in life, but others have no clinical consequences (13, 41).

1.5.3 Infection rates in infants with CF screen positive, inconclusive diagnosis

Respiratory tract infections with bacterial organisms such as *Staphylococcus aureus* and *Pseudomonas aeruginosa* are common in children with cystic fibrosis and are a cause of deteriorating lung function (12).

Studies have shown that patients with CFSPID have a higher incidence of CF associated respiratory pathogens such as pseudomonas aeruginosa or staphylococcus aureus compared to non-CF population (11, 17, 18). In addition, more pseudomonas infection is noted in CFSPID patients who convert to CF (19).

One hypothesis is that these cases were acquired when CFSPID patients present in clinic and were potentially exposed to patients with CF who are colonised with these pathogens (11). It is also possible that they have been picked up due to better microbiological surveillance at specialist CF centres (34). This risk of acquiring these pathogens in CF centres should be reduced with appropriate segregation in clinics and the specialist care of CF clinicians (34, 55).

1.5.4 Psychological impact of a designation of CF screen positive, inconclusive diagnosis

It is recognised that an inconclusive designation of CFSPID is an unintended consequence of screening and may cause long-term psychosocial, medical, and financial impact (18). The inconclusive and delayed diagnosis causes uncertainty for

clinicians and families and counselling of patients can be difficult due to the uncertainty of the prognosis (3, 33).

A short diagnostic period is associated with fewer negative parental feelings and increased confidence in the medical profession (57). Furthermore, early diagnosis allows time for genetic counselling and for parents to make decisions regarding future pregnancies (34). Therefore a well-designed newborn screening programme has the ability to reduce the psychosocial effects of a delayed diagnosis (34).

A study in Brittany found that a third of couples with a child previously diagnosed by newborn screening decided to opt for a prenatal diagnosis of a subsequent pregnancy (47). However, the result of the screening process often has no impact on reproductive decision making (47).

Regarding CF carriers, there is a difference of opinion over their identification and the potential psychological effect this may have (58). Parents have been shown to attach value to newborn protocols that reduce the number of carriers detected; as if communication is not clear, parents may believe that their child who is a carrier is "not normal" and may be at a risk of becoming unwell and developing cystic fibrosis (34, 58). Recognising carriers can lead to difficult conversations and can provide information about the extended family which may be unwanted (59).

Although carriers are predominantly healthy, some certain conditions are seen in a higher prevalence such as idiopathic bronchiectasis, pulmonary non-tuberculous mycobacterial (NTM) infection, chronic rhinosinusitis, idiopathic chronic pancreatitis, infertility, and diabetes (34, 39). In addition, it is important to note that future pregnancies are not risk-free and there is a small risk, 1 in 200, that future children of the couple could be born with cystic fibrosis and therefore they may opt for genetic counselling (1).

There is further psychological impact to consider regarding the risk of conversion. Conversion suggests a change in health status and when considering a change from

CFSPID to CF this could imply that life expectancy may be significantly shortened if not eligible for modulator treatment (4). The long-term psychological effect of a designation of CFSPID should be considered and psychological outcomes can be improved by including a psychologist a part of the multidisciplinary team.

1.5.5 CFTR related disorder (CFTR-RD)

Infants with CFSPID are at a greater risk of developing a *CFTR* related disorder (*CFTR*-RD) later in life (3, 17). *CFTR*-RD is a single organ disease associated with *CFTR* dysfunction but does not fulfil criteria for CF diagnosis and cohorts have a normal sweat test (2, 17).

It is beneficial to consider CFTR-associated conditions on a spectrum from those who are asymptomatic to those with severe multi-organ disease (7). A diagnosis of *CFTR*-RD is made from clinical symptoms whereas CFSPID is the result of a positive newborn screening result regardless of presence of any clinical symptoms (13). However, CFSPID has been considered a CFTR-RD with the clinical feature being hypertrypsinogenaemia (7). Therefore, there is some debate over whether children with CFSPID who develop clinical features should be classified as *CFTR*-RD or CF and this depends on the individual case (13).

The most common *CFTR*-RD is the congenital bilateral absence of vas deferens (CBAVD), which is a cause of male fertility (59); however other conditions include pancreatitis, isolated bronchiectasis and rhinosinusitis (13). Men with CBAVD often have *CTFR* variants on genotype analysis which are frequently VVCCs (13), however it can also be caused by non-CF causing variants (7).

1.6 Objectives

This thesis will be split into three objectives that will be explored throughout.

1.6.1 Objective 1: CF screen positive, inconclusive diagnosis population

This objective considers the number of individuals with a potential designation of CFSPID present on the UK CF Registry, their genotype and any trends identified in numbers of CFSPID and NBS.

1.6.2 Objective 2: Conversion from CF screen positive, inconclusive diagnosis to a diagnosis of cystic fibrosis

Objective two explores the CFSPID population to determine whether any fit the three criteria for conversion to CF which are variants reclassified as being CF-causing, increased sweat chloride or clinical symptoms consistent with a diagnosis of CF.

1.6.3 Objective 3: Sweat chloride analysis

The final objective considers the amount of repeat sweat test data for the CFSPID population which is essential to monitor for conversion to CF. The number of sweat tests will be evaluated alongside age at sweat tests, time between the diagnostic sweat test and first follow-up, and the trends in frequency of sweat test data recorded in the registry throughout the study period.

2. Methodology

2.1 Study design, setting, data sources and participants

We conducted a retrospective cohort study with participants from the UK CF registry.

2.1.1 Data Source

Data for this study was obtained with a data request form which was sent to the UK Cystic Fibrosis Trust. This form contained information on the research question, description and rationale of the project, the intended use of data and exact details of the data set required is present in the appendix.

NHS research ethics approval (Huntingdon Research Ethics Committee 07/Q0104/2) was granted for the collection of data into the UK database. The Cystic Fibrosis Trust database committee approved the use of anonymised data in this study.

2.1.2 Data Characteristics

Data was received in pseudonymised format. The data included baseline and annual review data. Baseline data requested included the variables: date of birth, gender, ethnicity, birth weight, date of diagnosis, all fields associated with the newborn screening result, screening lab, sweat test results and dates, genotype, faecal elastase result, social deprivation score. Annual review data included pancreatic status, weight, FEV1, Pseudomonas aeruginosa acquisition and number of hospital admissions. However, for this part of the study only certain variables from the baseline data were used and no variables from data on annual reviews were used.

2.1.3 Study Population

Although there were prior regional NBS programmes in Wales, Northern Ireland and certain parts of England, screening become universal across the United Kingdom in October 2007 (1). Therefore, data was requested for all individuals

added to the registry with a new diagnosis between 2007 and 2019. The first year of analysis was 2008 as it provided a full year of screening therefore would be comparable with the following years.

NBS is not the only way to diagnose CF as it is possible to suspect a diagnosis of CF antenatally on an ultrasound scan, just after birth due to meconium ileus or with the emergence of clinical symptoms and a failure to thrive. Only individuals diagnosed through NBS were eligible to be included in this study as CFSPID is an unwanted outcome of newborn screening, so individuals were not identified with a designation of CFSPID before the introduction of universal screening in 2007 (10). Individuals who were diagnosed through the other methods were excluded. To summarise, the study population was defined as children born between 2008-2019 and diagnosed with CF by newborn screening.

2.2 Objective 1: CFSPID Population

Data for the study population were analysed to determine if it contained any individuals with a potential diagnosis of CFSPID. A potential designation of CFSPID was defined using the European CF Society Newborn Screening Working Group (ECFS NSWG) definitions of CFSPID/CRMS (13) and are outlined below.

2.2.1 Definition 1

The first CFSPID definition was defined as a normal sweat chloride, <30mmol/L, and 2 CFTR variants with at least one with an unclear phenotypic consequence (13). Unclear phenotypic consequence may refer to variants of varying clinical consequence (VVCC) or variants of unknown significance (VUS) and to determine the classification of a variant, the CFTR2 database was used (26).

If a variant was not listed on the CFTR2 database, it was searched for on other databases such as the CFTR1 database (60), where case reports looking at the clinical consequences of a variant can be found which can help classify a variant. If their status remained unclear, they were named a VUS. To differentiate this group

from CF carriers, individuals need to have two CFTR variants whereas carrier of CF would only have one CFTR variant.

2.2.2 Definition 2

The second CFSPID definition was an intermediate sweat chloride (30-59mmol/L) and one or no CF-causing variant (13). The CFTR2 database was used as above to classify the variants as either CF-causing, VVCC or VUS.

2.2.3 Sweat Tests

Originally when considering sweat test data, only the first sweat test performed (ST1) was used to classify the individual. However, to maximise the detection of the CFSPID population and as further evaluation of the data showed that several individuals had a repeat sweat test within the first 31 days, all sweat tests within 31 days were considered as they were also deemed to have been used to assist diagnosis of these individuals. In addition, it was noted that a few individuals did not have a value for sweat test one (ST1) but then had subsequent values for sweat test two to five, hence the reason to include all sweat test results within 31 days of the first. The highest value was used to make a judgement on the classification of the sweat chloride

2.2.4 R117H Poly-T Haplotype

Individuals with at least one copy of R117H need the poly-T haplotype of this gene considered in more detail to determine whether it is CF-causing or a variant of varying clinical consequence.

The registry began collecting information of poly-T tract data in 2016 and records this data in three fields; 5T, 7T or 9T (8). A 5T haplotype is considered CF causing and 7T or 9T is considered a VVCC. Those with a CF-causing haplotype were removed from the CFSPID population if deemed to have CF and if an individual had a 7T or 9T haplotype they are considered to have CFSPID.

2.2.5 Reversed Diagnosis

There was a subgroup of patients in the study population had their diagnosis of CF reversed on the registry and this was recorded in a separate variable which provides information on the reason for reversal. These individuals were screened to ensure that any individuals who had their diagnosis reversed due to a designation of CFSPID were not overlooked due to missing genotype or sweat test data or as they are no longer considered part of the UK CF Registry.

The factors that can constitute a diagnosis of CF to being reversed include repeat normal sweat testing, CFSPID designation or carrier status. They were analysed considering their reason and age of reversal, genotype and sweat test data to determine if there are any CFSPID patients in this cohort.

The final CFSPID population consisted of results from both CFSPID definitions along with individuals from the diagnosis reversed subgroup deemed to be CFSPID to create a final CFSPID cohort.

2.2.6 Missing data

2.2.6.1 Genotype

An individual's genotype can be entered onto the registry with unclear terms such as missing, not known or not identified which can cause problems as it is not clear whether this suggests a rare variant of unknown significance (VUS) or whether the genotype data was simply omitted from the registry.

In addition, if a variant was not identified it is possible that the individual is a carrier for CF and therefore should not be present on the registry as they only have one CFTR variant. Therefore, individuals with missing or unknown data were not included in the CFSPID population but were be considered to determine whether inferences can be made regarding their classification as CF, CFSPID or a carrier.

2.2.6.2 Sweat test data

Patients may not have any sweat test data recorded on the registry which means that decisions cannot be made about their disease status. Individuals with no sweat test data were not included in the CFSPID population.

In addition, sweat chloride may have been overwritten so it was important to look at the date of the first sweat chloride to consider if this was the diagnostic sweat test. Age at first sweat test was evaluated to determine whether the age of the CFSPID population was older than the main study population which could be used to determine whether overwriting has occurred as it was not clear whether this is common practice with CFSPID patients.

2.2.7 Trends in CF screen positive, inconclusive diagnosis numbers

The CFSPID designation was introduced in 2014 and it was important to consider whether the publication of guidelines for CFSPID has impacted the number of children with potential CFSPID added to the UK CF registry.

First, the number of children diagnosed by NBS was compared alongside the number of children identified as a potential CFSPID designation for each year of the study (2008-2019). The amount of missing data for each year was evaluated to determine whether this may have led to an over or underestimation of CFSPID patients identified each year. Total numbers diagnosed by newborn screening each year was broken down into those with CF, CFSPID or missing data.

The study period was then split into two five-year periods from 2008-2013 and 2014-2019 to compare pre-CFSPID designation (2008-2013) to those identified after the designation (2014-2019). The numbers of CFSPID were compared in addition to the CFSPID population as a proportion of those diagnosed by NBS in each period.

2.3 Objective 2: Conversion to a diagnosis of CF

The next stage of the study was to consider those in the CFSPID population who had converted from a potential designation of CFSPID to a CF diagnosis. Individuals can convert to a diagnosis of CF due to changes in their clinical picture or the CF knowledge base. Criteria that indicate a change in disease status are reclassified variants, increased sweat chloride or the emergence of symptoms suggestive of CF.

2.3.1 Reclassified Variants

As more knowledge emerges on the clinical consequences of a variant its classification in the CFTR2 database can change. The CFTR2 database was used to look specifically for variants that have converted from either VVCC or VUS to CF-causing.

The date of conversion listed on the CFTR2 database was evaluated along with the birth year and diagnosis date of the individuals on the UK CF registry to determine whether the variants had been reclassified whilst they were identified as CFSPID or alternatively if the variant was CF-causing at the time of their diagnosis. Information on the reclassification of variants in the CFTR2 database is correct as of January 2021.

2.3.2 Sweat Chloride

Multiple sweat chloride results allow for the longitudinal analysis of disease status and can illustrate trends such as rising sweat chloride level over time. If a patient originally had a sweat chloride in the normal or intermediate range (30-59mmol/L) but a subsequent raised sweat chloride (>60mmol/L) they were considered to have converted to a diagnosis of CF. Available sweat test data after the first 31 days were assessed and any individuals with raised sweat chloride concentrations were identified. The date of conversion and genotype of these individuals was noted.

Sweat test results were analysed considering how close their value is to the diagnostic threshold, the date of the conversion and the genotype of the patient.

2.3.3 Clinical Symptoms

Another reason for a conversion is a clinical diagnosis of CF. If an individual had an unclear NBS result but then developed symptoms that fit with a diagnosis of CF this would be considered a conversion due to the emergence of clinical symptoms

It is possible for an individual to have more than one method of diagnosis selected on the CF registry. Therefore, a clinical method of diagnosis could have been selected on the CF registry alongside diagnosis by newborn screening for a single patient.

It is not clear if to what extent clinicians record additional methods of diagnoses over time and therefore it may not be possible to identify all individuals who have converted due to the commencement of clinical symptoms, which could lead to underestimation of those who have converted to CF.

In a data amendment, more information was received on method of diagnosis and date of diagnosis which can show if another method of diagnosis was added retrospectively overwriting the original diagnosis date. The CFSPID population was screened to consider those with an additional method of diagnosis other than newborn screening.

A full list of methods of diagnosis is present in the appendix and the methods that were indicative of a clinical diagnosis were selected and are present separately in the appendix. Individuals that had a diagnosis from the clinical diagnosis list were then investigated in more detail considering sweat test data and the time between birth date and diagnosis date to determine whether the additional diagnosis was made at birth or later.

2.3.4 Average time to conversion

The time to conversion was calculated considering the median time to conversion and evaluating the minimum and maximum time. There are inevitably differences in the populations, for example considering genotype, prevalence of different mutations and how often sweat tests are performed. The more frequently sweat tests are performed, the earlier conversions can be detected. Therefore, it is important to consider the time between the diagnostic and to the next sweat test in this conversion population.

Average time to conversion could help advise the appropriate length of follow-up for CFSPID patients. It could provide a benchmark age for clinicians to help inform decisions around discharge and when to be more vigilant with sweat tests.

2.4 Objective 3: Sweat chloride analysis

It is important that individuals in the CFSPID population are routinely followed up to show disease progression and to identify conversions; therefore, it was deemed important to investigate sweat chloride data stored in the registry in more detail.

The number of sweat tests in the CFSPID population was evaluated by calculating the total number of sweat tests each individual had and summarising this for the whole CFSPID population. In addition, the number of individuals with repeat sweat tests within 31 days and the number of individuals with at least one non-diagnostic sweat test (i.e., after 31 days) was calculated.

2.4.1 Age at first recorded sweat test

Next, age at each sweat test was evaluated for sweat test 1 through to sweat test 5. This will be useful to consider whether sweat tests have been overwritten for the CFSPID population and show the mean ages at which sweat tests are performed which shows if the guidelines are being adhered to.

2.4.2 Time to next sweat test

The average time between the diagnostic sweat test and first non-diagnostic sweat test after the first month was evaluated to determine whether CFSPID patients receive sweat test follow-up. Those with a repeat sweat test within 31 days were not included in this calculation to show a truer image of follow-up, not a repeat sweat test in the diagnostic period.

Current guidance suggests that CFSPID patients should have yearly sweat tests. This analysis evaluates whether clinicians are reviewing patients in line with the current guidance and to determine whether sweat test data from these reviews are consistently being added to the registry.

2.4.3 Consistency of follow up

To consider consistency of follow up, the proportion of individuals with at least one sweat test was calculated at each age (up to 5).

First, the population was split to consider the number of sweat tests before and after the introduction of the CFSPID designation in 2014. Next, all birth years since the introduction of the CFSPID designation (2014-2019) were considered to determine if there have been any changes more recently in terms of follow-up.

2.4.4 Sweat test follow up after 6 years

Finally, the length of follow up and the number of patients with follow up at age 6 and adolescence were evaluated as individuals are unlikely to convert after this age. Guidelines for CFSPID state that reviews of patients can be less frequent and can be tailored to the clinical picture of the patient. Numerous patients may be lost to follow-up or discharged by their clinician and live healthy lives and not require repeat sweat testing.

2.5 Statistical analysis

All analyses were carried out using R version 4.0.3 and R studio version 1.3.1093. More information on the statistical analysis plan, code and packages used are detailed in the appendices.

3. Results

3.1 Objective 1: CF screen positive, inconclusive diagnosis population

3.1.1 Participants

Data was requested from the UK CF registry for all new diagnoses in the period from 2007-2019 and the next step was to exclude individuals born before 2008 or those who were not diagnosed by newborn screening, which is shown in **Figure 3**. This resulted in a final study population of 2424 individuals who were born in or after 2008 and were diagnosed by newborn screening.

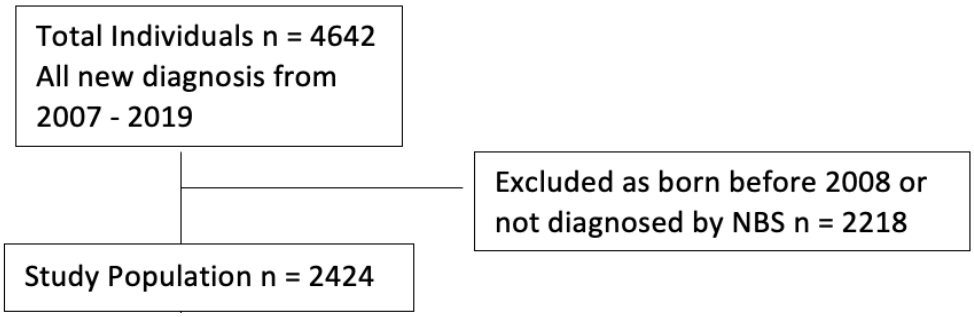


***Figure 3**: Identification of the study population excluding those not diagnosed by newborn screening or born before 2008*

3.1.2 Defining the CF screen positive, inconclusive diagnosis population

Out of the 2424 individuals in the study population, ten individuals had their diagnosis reversed and were considered separately to determine if their diagnosis reversal was due to a designation of CFSPID. Six individuals had missing data on their diagnostic status, however none were considered to potentially have a designation of CFSPID due to their genotype and sweat test data.

Therefore, as shown in **Figure 4**, 2408 individuals did not have their diagnosis reversed and thus had their sweat test and genotype considered to determine if they have a designation of CFSPID.

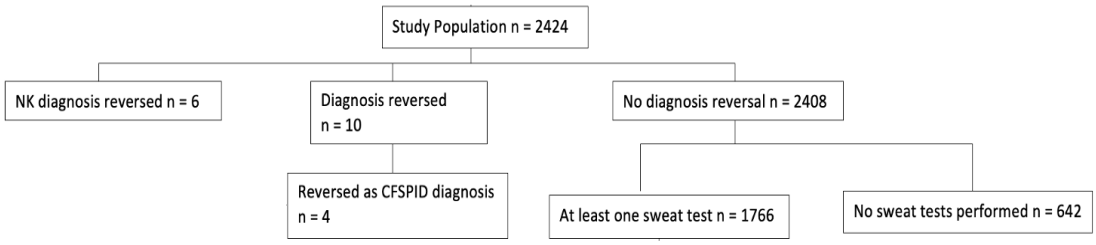


Figure 4: *Individuals on the UK CF Registry that had their original diagnosis reversed, those with no information on diagnosis reversal and sweat test data on those without a diagnosis reversal*

3.1.3 Sweat test data

To identify an individual as potentially having CFSPID, at least one sweat test is required as a normal or intermediate sweat chloride value makes up part of the definitions for CFSPID. Therefore, the study population was filtered for at least one sweat test and out of 2408 individuals, 1766 had at least one sweat test performed.

The individuals with at least one sweat test performed can be further classified into raised, intermediate and normal sweat chloride levels. A raised sweat chloride is classified as >60mmol/L, an intermediate sweat chloride is 30-59mmol/L and a normal sweat chloride is <30mmol/L. the number of individuals with each classification are shown below in **Figure 5**.

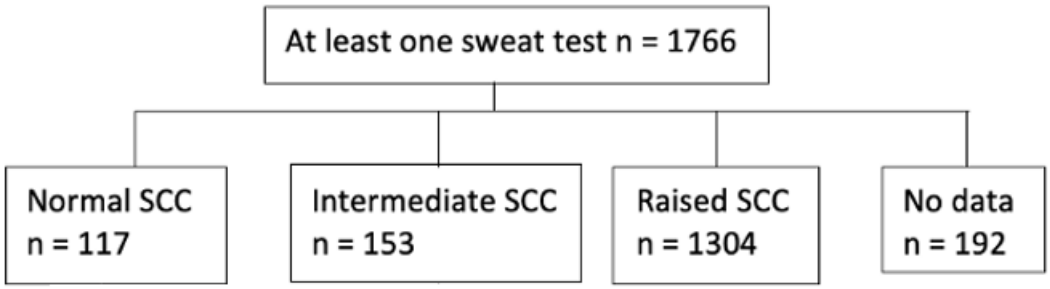


Figure 5*: Sweat chloride classification as normal, intermediate, raised or no data*

The use of multiple sweat tests within 31 days to identify individuals with a potential CFSPID designation resulted in two extra individuals being added to the CFSPID population which is shown below in **Table 1**.

	Only ST1	All sweat tests <31days
Normal sweat chloride/ CFSPID population	114/95	117/97
Intermediate sweat chloride/ CFSPID population	149/84	153/84
Total CFSPID population/ including diagnosis reversed	179/183	181/185

***Table 1**: Considering the timing of the diagnostic sweat test and how this affects the numbers of individuals in the CFSPID population*

3.1.4 Genotype Analysis

The next stage was to screen the genotypes of the individuals in the normal and intermediate sweat chloride populations to determine whether they fitted the CFSPID definitions set out by the ECFS NSWG (13).

3.1.4.1 Definition 1

There were 97 individuals in the normal sweat chloride group, that fulfilled the criteria for definition one, which was a normal sweat chloride and 2 CFTR variants with at least one with unclear phenotypic consequences.

Their average sweat chloride was 23 mmol/L, and the variants present in the population, both VVCC and VUS are listed in **Table 2**. There were more VVCCs present in the population than VUS with F508/R117H being the most prevalent variant in this population.

There were two individuals in this group with missing genotype information, so it was not possible to classify them as CF or CFSPID. They both were heterozygous for F508del with a missing second variant. Therefore, no decision was able to be made on the diagnostic status of these individuals. After removing these two patients, it resulted in 18 patients in the normal sweat chloride population with two CF-causing variants who did not qualify as CFSPID.

3.1.4.2 Definition 2

84 individuals in the intermediate sweat chloride group fulfilled the criteria of definition two, which required an intermediate sweat chloride and one or no CF-causing variants.

The average sweat chloride for this population was 40 mmol/L and the variants present in this population are also listed in **Table 2**. Comparing the variants found in the normal and intermediate sweat chloride, there was similarity with the proportions of VVCC's however there was a lot of variability between the two groups when considering the VUS's.

There were no individuals with missing genotype data in the intermediate sweat chloride population so all individuals could have their genotypes assessed for CFSPID. This resulted in 61 individuals with intermediate sweat chloride with at least one CF-causing variant.

Def 1 (normal SC, 2 CFTR variants)	Total: 97	Def 2 (intermediate SC, 0/1 CFTR variants)	Total: 84
VVCC		VVCC	
R117H	72	R117H	53
D1152H	18	D1152H	10
T1246I	<5	T1246I	<5
TG12TS	<5	Gln1291His	<5
3154T>G	<5	c.709C>G (Gln237Glu)	<5
		R1070W	<5
		5T; TG13	<5
		Y1032C	<5
		D443Y	<5
		F575Y	<5
VUS		VUS	
Cys491Phe	<5	2789+2insA	<5
M952I	<5	G502X	<5
c.3717+10Kbc	<5	L375F	<5
E1124del	<5	c.1384G>T	<5
		p. Ser118Phe	<5
		c.490-1G>C	<5
		c.4339del	<5
		c.473G>A	<5
		Y1073C	<5

***Table 2**: Genotype of individuals in CFSPID population outlining the number of VVCC and VUS present in both the normal and intermediate sweat chloride populations*

Variants that are present in less than five individuals in the CFSPID population are shown as <5 to protect their anonymity due to the possibility of re-identification.

3.1.5 R117H Poly-T Haplotypes

It was important to consider the R117H variant in-depth as the 5T haplotype would be considered CF-causing. Therefore, all those in the CFSPID population with the R117H genotype were explored further to exclude those with CF rather than CFSPID.

In total, there were 11 individuals with at least one copy of R117H;5T. Out of these 11 individuals, 8 were considered to have a diagnosis of CF and were excluded from the CFSPID population which is shown in **Figure 6**. All 8 had a sweat test in the intermediate range with an average sweat chloride concentration of 50mmol/L. This is a higher mean sweat chloride than the CFSPID population of 31mmol/L which is more suggestive of a CF-causing variant.

There is a lot of missing R117H poly-T haplotype data with only 49% of the CFSPID population having a complete dataset. It is possible that there are individuals incorrectly given a CFSPID designation with a 5T haplotype who were not detected due to missing data. This could have led to an overestimation of the potential CFSPID population.

Combining the 97 patients for definition 1 with 84 patients for definition 2 and 4 who had their diagnosis reversed results in a total for the CFSPID population of 185 individuals shown in **Figure 6**.

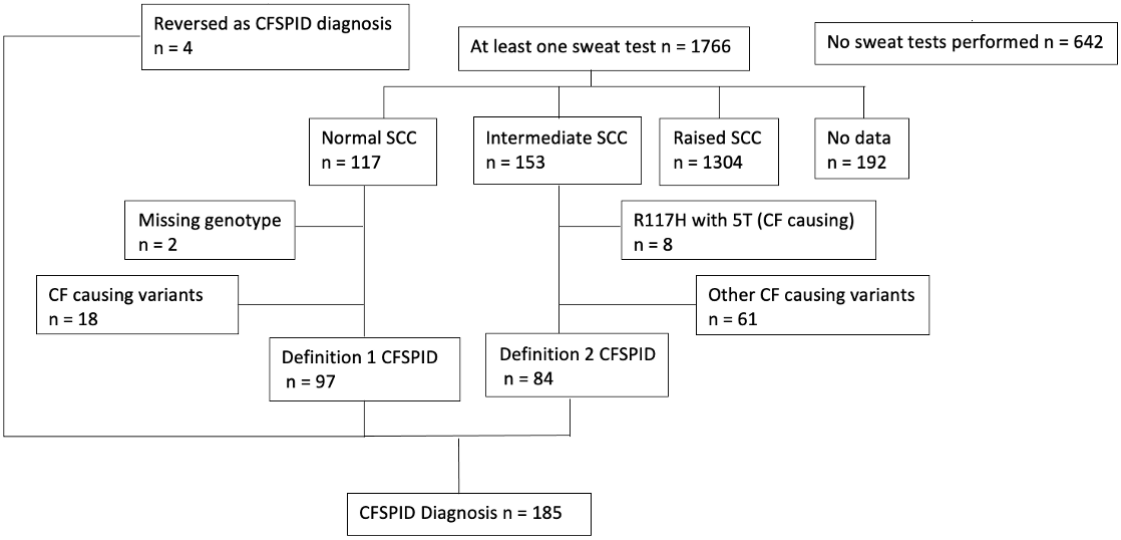


***Figure 6**: CFSPID diagnosis population diagram identifying those with a potential CFSPID diagnosis from both the normal and intermediate sweat chloride groups*

3.1.6 Trends in CF screen positive, inconclusive diagnosis and newborn screening data

3.1.6.1 CFSPID numbers

Considering the numbers of individuals with CFSPID in the two time periods, there are potentially 86 individuals with a designation of CFSPID in the first period from 2008-2013. In the second period from 2014-2019, 99 individuals potentially have CFSPID. Therefore, there has been an increase in the number of CFSPID designations after the introduction of the CFSPID designation in 2014 which is visible in **Table 3** and **Figure 7**.

3.1.6.2 CFSPID as a percentage of NBS

It is important to consider these numbers in the context of the total number of those with positive newborn screening in that period. There were 1251 individuals with a positive NBS in 2008-2013 however this number was lower in the next period, 2014-2019 with 1157 individuals considered. This results in 6.9% of positive newborn screening being potential CFSPID in 2008-2013 and 8.6% in 2014-2019. Therefore, there is no evidence of a reduction in the number of CFSPID identified after the introduction of the CFSPID designation in 2014.

3.1.6.3 Missing sweat test data

There is a considerable amount of missing sweat test data in the early years which could result in the CFSPID population being underestimated in the first few years as it was not possible to classify individuals as CFSPID without sweat test data. The percentage of missing data is present in **Table 3** and shown graphically in **Figure 7**.

3.1.6.4 Newborn screening trends

The number of children diagnosed by newborn screening has been decreasing in the last couple of years (2018-2019) according to registry data.

Year	Number of NBS	Number of CFSPID	% Of NBS with CFSPID designation	Number of missing Data	% Of NBS with missing data
2008	182	8	4.4%	106	58.2%
2009	192	8	4.2%	98	51.0%
2010	253	17	6.7%	116	45.8%
2011	214	22	10.3%	75	35.0%
2012	218	18	8.3%	50	22.9%
2013	192	13	6.8%	56	29.2%
2014	192	17	8.9%	47	24.5%
2015	203	18	8.9%	42	20.7%
2016	225	18	8.0%	49	20.9%
2017	206	13	6.3%	64	30.6%
2018	178	20	11.2%	34	18.5%
2019	153	13	8.5%	55	34.6%
2008-2013	1251	86	6.9%	501	40.4%
2014-2019	1157	99	8.6%	291	24.5%

Table 3: *Number of individuals with a positive NBS result, potential CFSPID designation and missing data for the years 2008-2019 and numbers when split into the time periods before and after the introduction of the CFSPID designation in 2014*

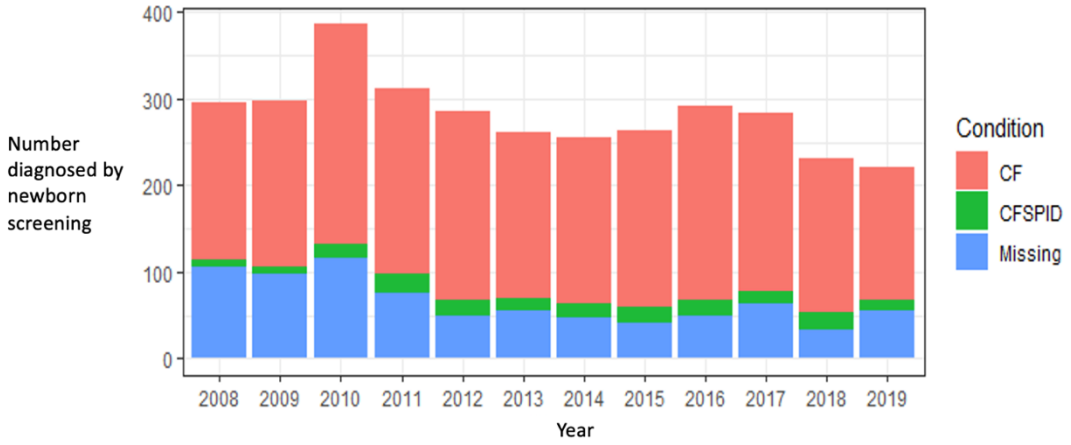


Figure 7: *Number of individuals with a positive NBS result, potential CFSPID designation and missing data for the years 2008-2019 and numbers when split into the time periods before and after the introduction of the CFSPID designation in 2014*

3.1.7 Reversed diagnosis of cystic fibrosis

As mentioned above, there are 16 patients on the registry who either had their diagnosis reversed or have missing data concerning the reversal of their diagnosis. In total, 10 were confirmed to have had a reversed diagnosis of CF and when analysing the reason for the reversal, 4 (40%) were reversed due to a designation of CFSPID.

CFSPID designation reversal group had very little sweat test data with 1/4 having sweat test data added to the registry. In addition, all individuals had common VVCCs in combination with F508 as their genotype.

The individuals who had their diagnosis reversed for reasons other than CFSPID seem to have been removed due to only one *CFTR* variant being identified making them carriers. This group had more sweat test data than the CFSPID reversals with an average sweat test value of 30mmol/L which is similar to the CFSPID average of 31mmol/L.

3.1.7.1 Time to reversal

The median time to reversal of CF diagnosis is 2.5 years, minimum of 0 years and a maximum of 10 years.

3.1.8 Missing data

642 individuals had no sweat test performed and were considered separately to see if any individuals had common VVCCs (R117H and D1152H) can cause CFSPID. When considering the two genotypes that are the most prevalent in the CFSPID population, 21 individuals in this population, without a sweat test performed, had at least one copy of R117H and 5 individuals had at least one copy of D1152H and may therefore fit a designation of CFSPID.

There are also 192 individuals, where it is indicated that a sweat test has been performed but there is missing data which result in only the date of the sweat test being recorded not the value or both pieces of information can be missing. There were 24 individuals with at least one copy of R117H and one individual with at least one copy of D1152H. These are potential CFSPID patients that have missed due to missing data.

3.2 Objective 2: Conversion from CF screen positive, inconclusive diagnosis to a diagnosis of cystic fibrosis

There were 9 conversions in the CFSPID population due to an increase in SC, less than 5 conversions due to clinical symptoms and no conversions due to reclassified variants. The exact number of clinical conversions is not stated due to anonymity on the request of the UK CF registry team.

Considering the 9 conversions due to an increase in sweat chloride, 6 had straightforward conversions, 2 had transient conversions and 1 had a conversion with two samples being taken on the same day. Out of the 9 individuals converting to a diagnosis of CF, 2/9 had a normal sweat chloride concentration on sweat test one and the rest were in the intermediate range.

The average sweat chloride values for sweat test one to three and shown in **Table 4** and show an increase through the tests. The median time to conversion was 3-years 1-month (37 months) with a minimum time to conversion of 2 months and a maximum time of just under 10 years.

Sweat test number	Average sweat chloride concentration
Sweat test one	42 mmol/L
Sweat test two	60 mmol/L
Sweat test three	74 mmol/L

Table 4: *Overview of the average sweat chloride concentration for each sweat test in those who converted from CFSPID to CF due to a raised sweat chloride*

3.2.1 Genotype

Genotypes were considered to determine if any variants appear to commonly convert to a CFSPID designation. Over half of the conversions had at least one copy of the VVCC, R117H. Another three of the conversions were heterozygous for F508. These are some of the most common *CFTR* variants, therefore this does not suggest that they are more likely to be involved in conversions. They are just more prevalent in the CFSPID population.

3.2.2 Repeat Sweat Tests

As conversion looks at an increase in sweat chloride over time, individuals need to have at least one repeat sweat chloride measurement. 89 (48%) individuals in the CFSPID population had data from at least one repeat sweat chloride test which allows for the evaluation of potential conversion to CF.

3.2.3 Transient Conversion

Two individuals have originally been considered as a conversion due to a second sweat test being greater than 60 mmol/L. However, they both had a third sweat test which was in the normal/intermediate range. This leads to issues concerning whether this is a true conversion and whether these individuals should be managed as CF or CFSPID.

3.2.4 Reclassification of variants

There are two variants in the dataset that have had their CF status reclassified in the CFTR2 database, Y569D and H199Y. Y569D was present 4 times in the study population and H199Y was present once in the study population. However, they were not eligible for the CFSPID population as they either had no recorded sweat chloride data or had a sweat chloride concentration consistent with a diagnosis of CF.

3.2.5 Conversion due to clinical symptoms

Table 5 details all individuals in the CFSPID population who had at least one method of diagnosis noted other than newborn screening. This list was then screened, removing those with another method of diagnosis not consistent with a clinical diagnosis.

Method of diagnosis	Number of Individuals
Genotype	52
Family history	13
Persistent acute respiratory infection	6
Meconium ileus	<5
Malnutrition	<5
Liver disease	<5
Electrolyte imbalance	<5
Prenatal	0
Nasal polyps	0
Oedema	0
Bronchiectasis	0
Pancreatitis	0
Fertility	0
Rectal prolapse	0
Steatorrhea	0

***Table 5**: Clinical methods of diagnosis selected alongside newborn screening for individuals in the CFSPID population*

Individuals with the family history field checked were excluded as a family history of CF is not sufficient for a diagnosis. In addition, the individuals with the genotype field selected were also excluded from the conversion group as the individuals in this group do not have a genotype consistent with a diagnosis of CF. Finally, those with meconium ileus were removed as this presents in the neonatal period and does not represent a clinical conversion later in life.

This left six individuals in the persistent acute infection category, less than five individuals with malnutrition, less than five individuals with electrolyte imbalance and less than five individuals with liver disease. There was also overlap between the categories with some having multiple clinical methods of diagnosis. The average number of sweat tests performed in these individuals and average sweat chloride concentration is shown in **Table 6**.

This resulted in eight individuals having converted to a true diagnosis of CF due to clinical symptoms. The next step was to consider the time between birth date and diagnosis date. The average diagnosis date was eight months and as the newborn screening protocol takes about six weeks to result in a diagnosis due to the second IRT measurement take at 21 days, it is possibly that a later diagnosis has been added.

It is also expected that clinical symptoms may take some time to develop, and any new diagnosis added greater than three months from birth date was considered to potentially have been overwritten and possibly a second diagnosis added to the registry. Therefore, individuals with this gap were considered to have potentially converted to a diagnosis of CF.

Method	Average number of ST	Average ST value
Persistent acute respiratory infection	3	39
Malnutrition	4	55
Liver disease	1	27
Electrolyte imbalance	1	11

Table 6: *Individuals identified in the CFSPID population who may have converted due to clinical symptoms, the average number of sweat tests performed, and the average sweat chloride concentration*

3.2.6 R117H;5T Haplotype

One individual was originally considered a conversion due to an increase in sweat chloride from 56mmol/L to 84mmol/L over a 6-month period. However, on further evaluation, this individual had an R117H;5T genotype which is CF-causing, therefore they do not have a designation of CFSPID. They were removed from the conversion population.

3.3 Objective 3: Sweat chloride analysis

3.3.1 CF screen positive, inconclusive diagnosis longitudinal data

Just under half of the individuals in CFSPID population had only one sweat test performed and there was a rapid drop off in the number of sweat tests completed with 86% of individuals having two or fewer sweat tests, which is illustrated in **Figure 8** and **Table 7**. **Table 8** details 28 individuals (15%) had multiple sweat tests within the first 31 days which could be useful in cases of borderline sweat chloride. Furthermore, after these 31-days, only 64 (34%) had a repeat sweat test.

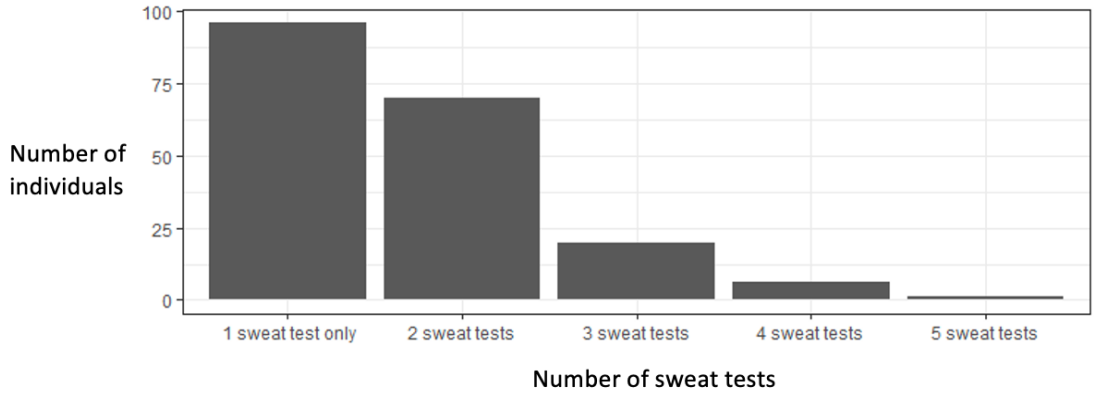


Figure 8: *Number of sweat tests performed in the potential CFSPID population*

Only 1 recorded sweat test	2 recorded sweat tests	3 recorded sweat tests	4 recorded sweat tests	5 recorded sweat tests
90 (49%)	66 (36%)	19 (10%)	6 (3%)	1 (0.5%)

Table 7*: Number of sweat tests performed in potential CFSPID population*

Multiple sweat tests in the first 31d	At least 1 sweat test after the first 31d
28 (15%)	64 (35%)

Table 8: *Number of sweat tests split before and after the diagnostic period*

3.3.2 Average sweat test value

The mean sweat chloride for sweat test one was 31mmol/L, which is just above the cut-off for the normal level. **Figure 9** below shows the distribution of the sweat test 1 values with many in the 20-39mmol/L range.

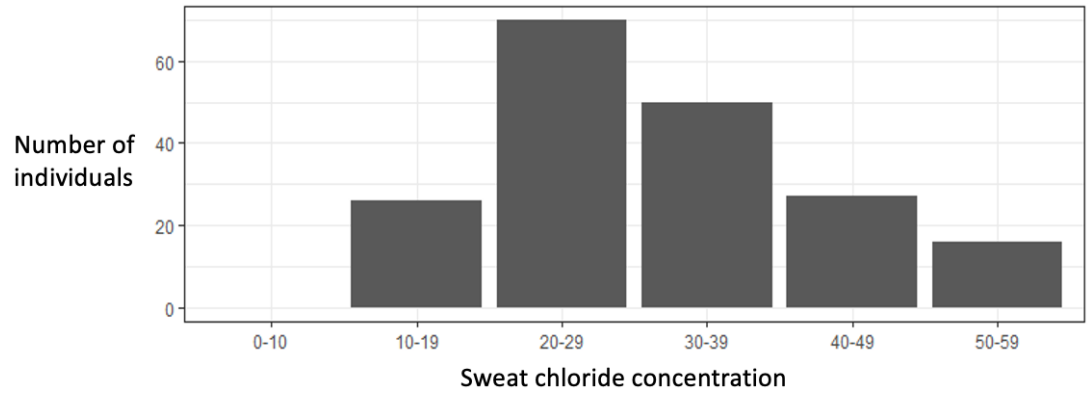


***Figure 9**: Average sweat chloride concentration at first sweat test in the CFSPID population*

3.3.3 Age at first recorded and subsequent sweat tests

Age at first recorded sweat test was useful to ensure that sweat tests had not been overwritten in the CFSPID population.

The mean age at the first sweat test was calculated in the CFSPID population as 91 days. This is in accordance with the NICE guideline which puts the median CF diagnosis age at 60 days (61). However, it is possible that there is some overwriting of sweat tests which has increased the mean age at first sweat test. **Table 9** shows the cumulative total of sweat tests in the population at each age.

Age in years	Sweat test 1	Sweat test 2	Sweat test 3	Sweat test 4	Sweat test 5
< 1	165	47	6	2	0
< 2	177	69	10	4	0
< 3	180	74	12	5	0
< 4	181	79	16	6	0
< 5	182	83	20	6	0
> = 5	182	92	26	7	1
Total (excl DR)	182	92	26	7	1

***Table 9**: Cumulative number of sweat tests performed at each age for all the CFSPID population from sweat test one to sweat test five*

3.3.4 Time to between diagnostic and first non-diagnostic sweat test

Time to next sweat test was considered for the CFSPID population as a relative indicator of how frequently individuals are followed up and have their disease status analysed and whether the ECFS guidelines of follow-up after six months are implemented. This was displayed in a **Figure 10** to show the range of follow up in the CFSPID population, the red line indicates the median time to the next sweat test. However, when constructing figure 5, repeat sweat tests performed in the first month were excluded to give a truer representation of time to follow up.

Figure 10 indicates that the majority of repeat sweat tests are conducted in the first year of life, with a median time of just over a year. However, it also shows a long tail demonstrating that some individuals are not followed up until later in childhood or previous tests have not been recorded.

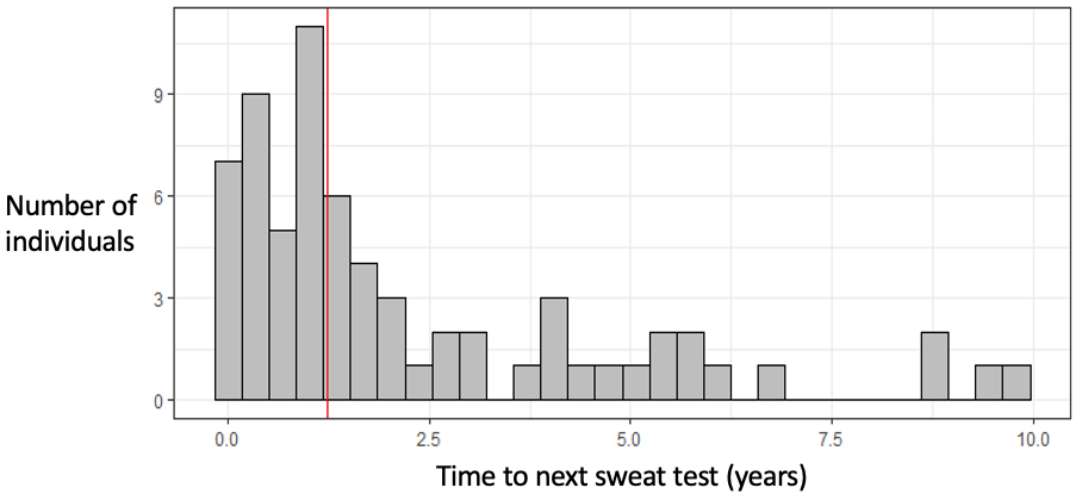


Figure 10: *Time between diagnostic and first non-diagnostic sweat test*

3.3.5 Consistency of follow-up

The proportion of individuals with at least one sweat test at each age is shown in the two graphs below. **Figure 11** splits the study period into two sections looking at before and after the introduction of the CFSPID designation in 2014. **Figure 12** looks at individual years after the introduction of the CFSPID designation in more detail to determine any more recent trends.

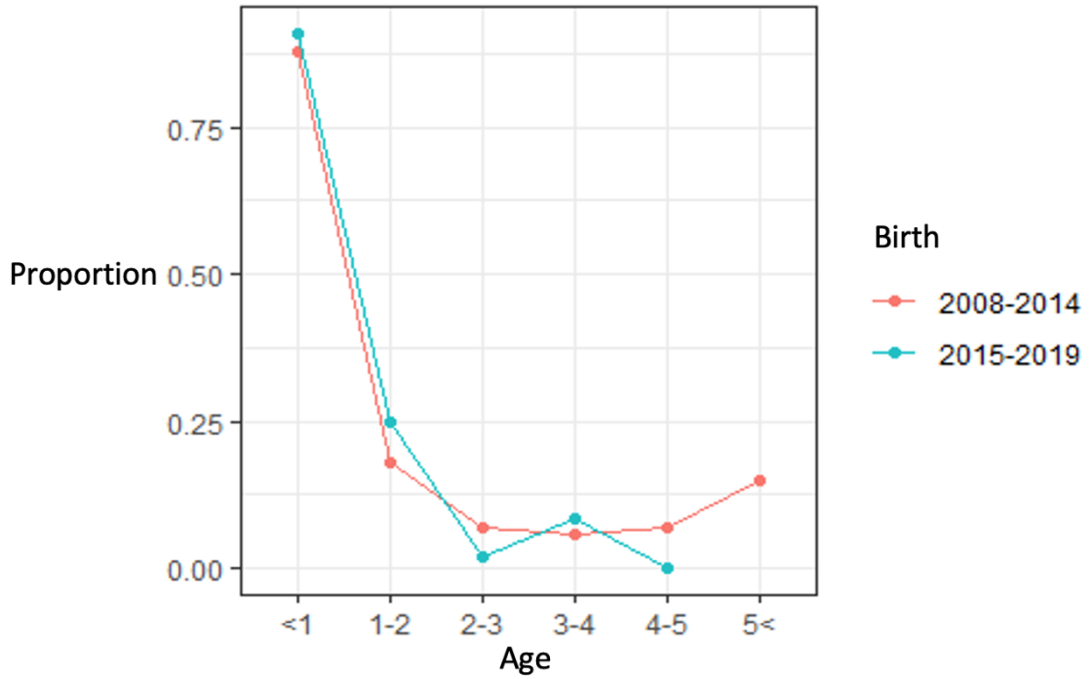


Figure 11: *Proportion of individuals in the CFSPID population with at least one sweat test at each age split before and after the introduction of the CFSPID designation in 2014*

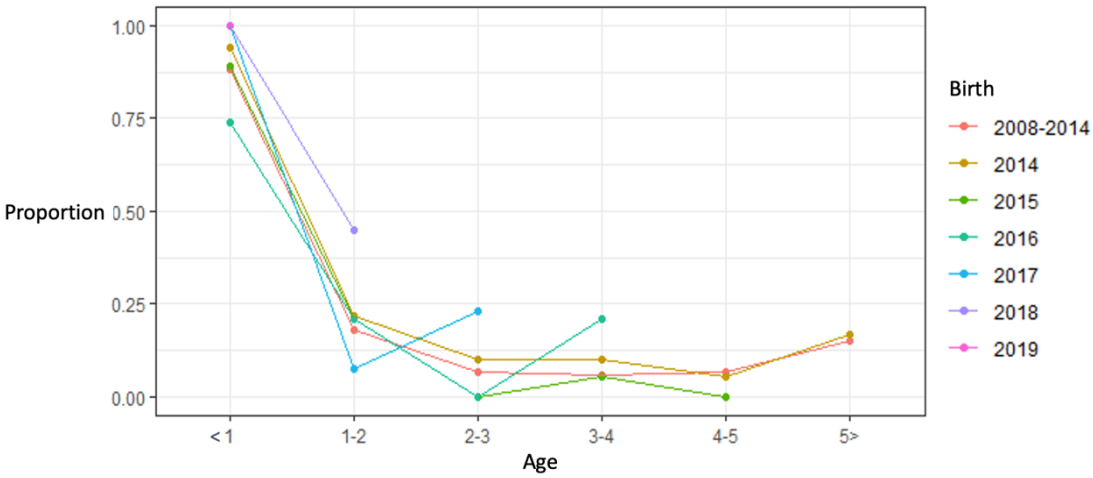


Figure 12: *Proportion of individuals in the CFSPID population with at least one sweat test split by year following the introduction of the CFSPID designation in 2014*

3.3.6 Sweat test follow up after age 6

Next, sweat test follow up after 6 years was analysed as individuals are unlikely to convert after this age and clinicians may not feel the need to continue seeing these patients annually. For this section of analysis six years of follow up was required so

only those born pre-2014 were included in this section of the analysis, which is 90 individuals, 46.6% of the CFSPID population.

Table 10 shows that only 10 (11.1%) of this subset of the CFSPID population had sweat test follow-up for greater than 6 years.

	CFSPID Population
Born pre-2014	86 (46.5%)
Sweat test follow up >6 years	13 (15.1%)

Table 10: *Number of individuals with at least six years follow up and those with a sweat test after six years of age in the CFSPID population*

4. Discussion

Since the introduction of NBS in the UK, a small but significant number of infants with a positive screening result have been included on the UK CF registry without clear evidence of a CF diagnosis. The characteristics identified suggest these infants have a potential CFSPID designation and the data available advises that evaluation on an individual level could be improved.

4.1 CF screen positive, inconclusive diagnosis population

4.1.1 CF screen positive, inconclusive diagnosis genotype analysis

4.1.1.1. Missing data

Out of the normal and intermediate sweat test groups which accounts for 270 individuals, only two had missing or inconclusive genotype information. There is very little missing genotype information in comparison to the significant amount of missing sweat test data.

Genetic testing has advanced in the last decade to become an integral part of CF diagnosis and the extensive variant databases allow for accurate classification of patients as CF, CFSPID or a CFTR carrier. In 2019, the UK CF registry reported that 99.2% of individuals were genotyped (62). Clinicians, when diagnosing individuals, may place more weight on their genotype information rather than sweat test data which may explain the completeness of genotype data in contrast to sweat test data.

Considering the two individuals with missing data, their disease status is not clear. This is because there is no clear option for 'no variant found' when entering genotype information. The information is instead added as 'miss/NK' which is vague as it is not clear whether no variant has been found or if the data is missing. A standard 'no variant found' variable would remove the ambiguity surrounding those without two *CFTR* variants.

4.1.1.2 Genotype recording on the registry

The input of variants into the registry was not consistent; the majority of individuals had their legacy name entered however, a few have their protein name or cDNA name entered onto the registry. This makes it difficult to quickly calculate the total number of each variant as the same variant can be entered in multiple ways onto the registry. A uniform way of referring to and inputting variants onto the registry would substantially reduce the amount of time it takes to analyse genotypes on the registry. This may not be possible for all variants as it seems that not all variants have legacy names that are commonly used, and the cDNA or protein name may be seen to provide more information on the variant making it easier to comprehend.

Adding multiple fields to the registry for all types of variant names would provide more information to those accessing the registry and if this could be done in an automated way it would not significantly increase the workload for those who have to input data to the registry. However, it is also fair to argue that this would just be a duplication of genotype information on the registry without providing any new information.

4.1.1.3 Variant classification

The presence of the CFTR2 classification of a variant on the registry would have assisted this study greatly and would have allowed a variant to be quickly identified as CF causing, VVCC, VUS or non-CF causing (26). This would aid those who do not have a great depth of knowledge beyond the most prevalent variants to be able to quickly classify and make inferences on the predicted clinical picture of these patients.

There are difficulties with this proposal as variants, especially those that are rare, may be reclassified as more individuals with this variant are documented and their clinical picture evaluated. This would result in the registry classification needing to be frequently updated to reflect the classification changes made in CFTR2 which increases maintenance work for the registry team.

If this is not feasible, then a central list of variants within the registry, highlighting especially those variants considered to be VVCCs and VUSs would provide a consistent set of variants to be aware of when evaluating for CFSPID. A pop-up box that indicates that a variant is a VVCC or VUS which may be more in line with a CFSPID designation than CF may result in fewer individuals with CFSPID being added to the registry.

4.1.1.4 CF causing variants with low sweat chloride

The Guidelines for the Performance of the Sweat Test for the Investigation of Cystic Fibrosis in the UK states that it is rare to have a normal sweat test along with a CF genotype (63); however, this study found several individuals with two CF-causing variants, but a normal or intermediate sweat chloride. Individuals with these results were present in both categories of sweat tests but there were significantly more individuals in the intermediate sweat chloride category.

In the normal sweat chloride group (<30mmol/L), 18 (15%) individuals had two CF causing variants; these patients should be considered to have CF due to the two CF causing variants even though their sweat chloride is below the diagnostic level. These individuals would not qualify for modulator therapy due to their lower sweat chloride unless they had clinical phenotype of CF. This group of patients may have a less severe phenotype than those with a sweat chloride diagnostic of CF and be pancreatic sufficient due to an incomplete molecular defect.

For those infants with an intermediate sweat chloride, 69 (45%) individuals had two CF causing variants and hence a clear CF diagnosis. It was expected that there would be more individuals in the intermediate range with this genotype and these individuals could have a sweat chloride close to the cut off for CF. As there is a margin of error with sweat tests, those close to the borderline may have a repeat sweat test which is then found to be diagnostic of CF. Individuals with R117H with a 5T haplotype are often found in this group, but have a diagnosis of CF, as R117H/5T is classified as CF causing (in contrast to R117H/7T, which is a VVCC). Eight

individuals with R117H/5T were identified in the intermediate sweat chloride group and as the registry only starting recording R117H poly-T haplotype data in 2016 this number could be much greater (8).

4.1.1.5 Proportion of Variants

When considering the genotype of individuals in both the normal and intermediate sweat chloride group, there are a similar number of variants of varying clinical consequence (VVCC) in both populations. The variants R117H, D1152H and T1246I are present in both the normal sweat chloride and intermediate sweat chloride groups and in similar proportions.

R117H was the most prevalent variant in both populations; accounting for 74% of the normal population and 66% of the intermediate. D1152H was the second most prevalent genotype accounting for 18.6% of the normal population and 10.9% of the intermediate. The third commonest variant was T1246I (1.0% of the normal population and 2.2% of the intermediate).

Concerning variants of unknown significance (VUS's), there are no particularly common variants in this group. This can make it difficult to predict which variants will convert to a CF diagnosis, as the variants are so rare and may only be present in a handful of individuals worldwide.

For non-CF causing variants, we did not recognise any in our potential CFSPID population. Although clarified as 'non-CF causing', they have been reported to be associated with CFTR-RDs in other papers.

4.1.1.5.1 VVCC

In the CFSPID population, 92% of individuals fit into the VVCC category with 8% of individuals having a VUS. Common VVCCs could be more likely to be added to the registry rather than rare variant as clinicians may recognise the VVCCs. They may not feel as confident to predict the phenotype of the rare variants.

4.1.1.5.2 VUS

When comparing the normal to the intermediate sweat chloride group there are slightly more VUS in the intermediate group.

4.1.1.6 Overall characteristics of potential CFSPID patients

The CF population is considered to have great heterogeneity with some having a mild clinical presentation and others having a much more severe phenotype. Even when considering individuals with the same genotype there can be phenotypic variance. The CFSPID population in contrast are by definition well children without symptoms.

In terms of the genotype, 70% are F508del/R117H and another 14% are F508del/D1152H which could suggest this is a homogenous group as nearly 85% have two genotypes. The next 15% all consist of individuals with a much rarer genotype.

4.1.2 The impact of the poly-T genotype on the R117H variant

The R117H variant can exist with three haplotypes: 5T, 7T and 9T, which reflect the length of the poly-T tract before exon 9. The shorter length of the poly-T region impacts the splicing of intron 8 and the expression of *CFTR* messenger RNA. The poly-T haplotype was considered when determining the classification of the genotype as 5T is considered CF causing whereas 7T is a variant of varying clinical consequence (VVCC).

Individuals with R117H;5T were found in the potential CFSPID population and had a sweat chloride in the intermediate range; this was expected as CF causing variants are more likely to cause a higher sweat chloride. These individuals were then removed from the CFSPID population.

A significant issue with the R117H group was the amount of missing data and this means the potential CFSPID population may contain individuals who have a 5T, CF causing haplotype, however, this information was not entered onto the registry.

Clinicians and registry staff have previously stated that R117H genotype data can be difficult to interpret, and genetic lab reports are often not clear which can impede the diagnostic process. In addition, there were difficulties with this classification and clarification had to be sought as quite a few individuals had 5T, 7T or 9T not on the R117H variant but the F508del variant.

Although the results may be difficult to interpret, all individuals with R117H should have their poly-T haplotype added to the registry to better characterise these individuals as CF or CFSPID. The registry is working to improve the amount of R117H poly-T haplotype data it records, and it is hoped that this information will more readily available in the future.

4.1.3 CFTR2 Database

The CFTR2 database was used in this study to classify variants, however, a handful of variants were not present on CFTR2 (26). These variants are rare and of unknown significance (VUS's). In these individuals, sweat chloride measurement is particularly important as there is very limited information available on these variants.

Some variants, not present in CFTR2 were included in the older CFTR1 database (60). The CFTR2 database is in the process of being updated and new variants are added subsequent to discussion with a panel of clinicians. Therefore, the variants not yet present on the CFTR2 database may be in the process of being added.

Variants on the CFTR1 database are often rare and a few individual patients provide information on phenotype (28). In addition, the CFTR1 database does contain variants that are heterogeneous in their clinical impact and not all are related to CF(28).

It is not possible to classify a variant based on only one case study especially as the variant could be a variant of varying clinical consequence and therefore can present differently in different individuals. This could make classifying these rare variants impossible until there is a significant number of case studies to use for comparison.

The CFTR1 database contained 6/13 of the VUS not present on the CFTR2 database. Clinical details on these six individuals are outlined in the appendix (compared to CFTR1 data). The overall clinical picture was similar in our cases to those previously reported but complicated by co-expression with VVCC's.

4.1.4 Trends in CF screen positive, inconclusive diagnosis numbers before and after the introduction of the CFSPID designation in 2014

It was hypothesised that the number of potential CFSPID cases would have decreased due to a greater awareness of what constitutes a potential designation of CFSPID and the increased knowledge that these individuals should not be placed on the CF registry as they do not have a diagnosis of CF.

The number of potential CFSPID cases on the registry seem to have increased following the publication of the CFSPID designation in 2014, however, the number of cases may have been underestimated in the earlier years of the study. There were substantially more missing sweat test data in the earlier years of the study and decisions on disease status are not possible for these patients. This makes it difficult to determine whether there has been an increase in CFSPID individuals or whether the recording of sweat test information has improved. It seems that greater education of CF clinicians is required on what constitutes a CFSPID designation and that these individuals require routine follow up.

The number of individuals that potentially have CFSPID that are not present on the UK CF Registry is unknown. It is not known how often these individuals have been followed up and whether any repeat sweat tests have been performed. The number of individuals with CFSPID not on the registry is likely low, but it would be difficult to calculate this as it would require data from every CF centre. This will be

affected by CF centre variance in knowledge and understanding of CFSPID which may result in certain centres adding more individuals with CFSPID to the UK CF registry. It is also possible that individuals with CFSPID are seen outside the CF service.

Individuals with CFSPID represent a new population for CF clinicians created as an outcome of newborn screening. There was no published guidance in the early years of this study and some clinicians have a lower default for following individuals up following a positive newborn screening result.

Considering the characteristics of individuals with CFSPID added to the registry, the vast majority of children on the registry have two identified *CFTR* variants and it is very rare for individuals to have one variant associated with carrier status. Therefore, two *CFTR* variants being identified, regardless of the phenotype they cause, seems to be sufficient to prompt a clinician to add a child to the CF registry. CF centres receive funding for adding individuals to the CF registry and this could influence centres to add children to the registry if they are unclear of the child's disease status.

When considering the missing data in the earlier years of this study, it is not possible to determine whether data is missing due to clinical visits not being recorded or whether these individuals did not receive any sweat test follow-up after the first diagnostic encounter. However, you would expect individuals to have at least one sweat test recorded as they had an initial positive newborn screening result (raised IRT).

The registry data set does show improvements from the data considered in 2008 to the current data in 2019/20. Sweat test data are more frequently available for these patients and there has been more follow-up data in the last couple of years.

As this is a large dataset with data recorded by clinicians or admin staff from CF centres across the UK, errors and missing data are to be expected. The registry does

have some internal data validation to verify the data being added. For example, sweat chloride concentrations can only be entered within a sensible range to avoid errors. Therefore, certain themes are visible in the data set and there certainly are potential CFSPID patients who have been added to the registry and then not adequately followed up.

4.1.5 Trends in newborn screening

The number of children diagnosed by newborn screening have decreased slightly in the last few years of the study. This trend has also been noticed in several other regions offering newborn screening to detect cystic fibrosis (7).

The decrease in the incidence of CF could be due to several reasons and one reason noted in European populations is the increase in prenatal diagnosis (7). Increased cascade screening in families that are at high risk for CF in the last few years can result in babies with CF being detected early in pregnancy which results in the option to terminate the pregnancy (16).

In the future, a much more significant percentage of CF could be diagnosed through cascade testing and antenatal methods. It is important to note that although the incidence of CF may be decreasing, prevalence is increasing due to modulator therapy increasing survival and life expectancy (55).

4.1.6 Diagnosis reversal

Some individuals who were originally on the registry had their diagnosis reversed for several reasons including a CFSPID designation. Every patient with CFSPID should theoretically have their diagnosis reversed on the CF registry and this could provide a solution and place to record these individuals on the CF registry without considering them a case of CF. Their newborn screening result and subsequent sweat tests would be recorded in preparation for a new CFSPID sub-registry.

4.1.6.1 Genotype

The genotype of the diagnosis reversed group seems to mirror the distribution of genotypes in the total CFSPID population as 75% have an R117H variant whereas the other 25% have a D1152H variant.

4.1.6.2 Sweat Testing

When considering sweat test data for this group, 25% had a recorded sweat test which was in the normal range. It is not possible to determine whether a sweat test was performed in the other individuals and if this aided the identification of CFSPID or whether the diagnosis was made solely due to genotype and possibly an absence of symptoms. Some individuals in this group had only one CF-causing variant which suggests they are *CFTR* carriers. This explains their diagnosis reversal as they do not have a diagnosis of CF. Determining what triggered the reversal of the diagnosis and when this occurred is important to gain an insight into clinicians' rationale.

4.1.6.3 Repeat Sweat Testing

However, CFSPID is not the only reason for a diagnosis to be reversed as two individuals had their CF diagnosis reversed due to repeat normal sweat testing. The genotype of these individuals was analysed as they may potentially be CFSPID but on further examination, they both had one CF-causing variant and one not known (NK) variant.

This group has a significant amount of missing data and a considerable number of individuals do not have any sweat test data. They likely had at least one sweat test performed to have their diagnosis reversed however it has not been entered on the registry. Sweat test data would have been helpful especially for the CFSPID population to better classify as definition 1 (normal SC) or definition 2 (intermediate SC) as per ECFS guidelines (13).

4.1.6.4 Average time to reversal

The average time to reversal was 2.5 years. With all the individuals in this group, it is important to understand why clinicians have been concerned with reversing the

diagnosis for these specific patients on the registry and why it is not done universally.

For carriers, it is understandable that they are removed from the registry as their diagnosis is clear and are not at risk of developing CF. There was possibly a delay in receiving the whole clinical picture that first had them added to the registry. When considering CFSPID patients removed from the registry, they had more typical genotypes of CFSPID so clinicians may have been more comfortable reversing their diagnosis but there are several others on the registry with a very similar genotype and sweat test results not reversed and it is not clear why they have remained on the registry. It seems that clinicians are very adept at only adding individuals with two CF variants to the registry, therefore there are only very few carriers present. However, they are not as good at identifying CFSPID on the registry, this may be due to anxiety over the conversion to CF or a gap in knowledge of what constitutes a CFSPID designation.

4.2 Conversion to a diagnosis of cystic fibrosis

4.2.1 R117H poly-T tract

Originally, 10 individuals were found to have converted to CF however, one individual was removed from this group once information on the R117H poly-T haplotype was received in the data amendment. This conversion was removed when their R117H poly-T haplotype was reported to be 5T, which is CF causing. Therefore, this individual is considered to have CF as their genotype is F508/R117H;5T.

The first sweat test performed in this individual resulted in a sweat chloride just below the diagnostic value of 60mmol/L and a subsequent sweat test within the first year of life demonstrated an increase into the CF range.

As the result above was not a conversion but originally a case of CF, the other individuals in the conversion group with an R117H allele should have their poly-T haplotype investigated in more detail. As the R117H poly-T haplotype was only added to the registry in 2016 this information may not be available for some individuals in this population (8).

However, if individuals have a 7T or 9T haplotype, these individuals can be deemed to potentially have CFSPID and an increase in sweat chloride above the diagnostic threshold can be considered a true conversion. However, if these individuals have an unknown R117H haplotype then it is not possible to state whether these individuals have a 5T, CF-causing haplotype or 7T or 9T, a VVCC genotype. Therefore, conversions could be overestimated in our population due to the lack of R117H poly-T haplotype data.

4.2.2 Repeat sweat test data

A barrier to determining conversions was a lack of repeat sweat test data. Only 50% of patients had repeat sweat test data recorded on the registry and routine sweat tests are required to identify a conversion, therefore the number of conversions could be significantly higher than identified. However, if these children were well and asymptomatic then there might not have been a clinical need for future sweat tests. Also, it is possible these sweat test were performed but have not been added to the registry. Nonetheless, recent guidelines require these individuals to be routinely followed up to monitor for conversion.

When considering the wider study population of 1766 individuals, only 368 (20%) had a repeat sweat chloride therefore our CFSPID population had a greater proportion of repeat sweat tests. This was expected as individuals with CF do not require repeat sweat tests unless they are monitored on modulator therapy. There were a few individuals in the CFSPID population who had an empty sweat test one (ST1) field in the registry but had data present in the sweat test two (ST2) field. This may also be present in the CF population; as it may be possible that their first sweat test was in the normal or intermediate range, therefore it was not added

to the registry. This would be considered a conversion, and this could mean that conversions may be missed due to the first sweat test data not being entered onto the registry. This is only speculation and the gap between sweat tests has to be considered to see if they are the same day or longer, but it would be informative to note the reason for the values for sweat test one to be missing.

This is not a concern if these individuals are receiving the correct treatment as CF patients and the fact that they may have originally had a designation of CFSPID is not pertinent clinically other than to be recorded correctly.

4.2.3 Borderline sweat chloride concentration

A borderline sweat chloride concentration is considered in this study to be between 50-60mmol/L. There was one individual in the conversion population who had a repeat sweat test the same day which saw a borderline sweat chloride increase to above the diagnostic threshold.

It is reasonable to repeat a sweat test if it is borderline as the categories are arbitrary and there is a margin of error in sweat tests. It is good practice to get a sweat sample from two limbs of the infant to ensure that a sufficient amount of sweat is collected for analysis which results in two sweat tests being recorded for an individual on the same day (63).

Discretion should be used to consider a diagnosis of CF if sweat chloride is within a range, for example, ±5mmol/L, if there are other reasons to suspect CF such as their genotype and clinical presentation. However, this does increase the clinical burden and that could lead to repeat sweat testing until a CF diagnosis. Furthermore, multiple sweat tests may increase anxiety for families and extend the diagnostic period.

4.2.4 Transient conversion due to increased sweat chloride concentration

This conversion population contained two conversions that were deemed transient. This is due to a conversion being noted on a second sweat test but then a subsequent third sweat test showed a normal or intermediate sweat chloride value. A group of sweat tests with this pattern results in a diagnostic conundrum where the disease status of an individual may be unknown as they have fit a diagnosis of both CF and CFSPID at different points of their life.

One option when considering the classification and management of these individuals is to treat them as if they have CF as at one point, they did have a sweat test in the diagnostic range, and it is unknown for how long they exhibited this high sweat chloride.

In contrast, they could also be considered CFSPID due to the two measurements in the normal or intermediate category and the high results could have been a one-off due to biological variation or an error in the accuracy of the sweat test. The proximity to the borders of the definition should be considered, however the management would have to be at the judgement of the clinician.

The situation would be simpler if only two sweat tests were performed however, this could lead to the overmedicalisation of healthy children with very expensive modulator therapy which is not a good use of healthcare resources. On the other hand, it could encourage repeat sweat tests until a clinician receives a sweat chloride value consistent with a CF diagnosis.

This results in a trade-off between fully treating every child with CF or preventing the overmedicalisation of healthy children. The clinical presentation of the individual has to be at the heart of this diagnostic decision making especially as the nuances of clinical tests render them challenging to interpret.

When considering this circumstance, it is important to consider sweat test values, the dates the tests were performed and the gaps between their tests.

4.2.5 Clinical symptoms leading to conversion

The presence of clinical symptoms can result in a conversion from CFSPID to CF and we identified less than five individuals in the CFSPID population who fit the criteria. It is unclear whether individuals commonly have a clinical diagnosis added which could lead to an underestimation of the conversions in this category. The original diagnosis method of NBS may have been overwritten to record the new diagnosis method of clinical symptoms. That would make it difficult to detect these individuals however the time between birth date and diagnosis date gives a good indication of whether another diagnosis method has been added subsequently.

As newborn screening can take around six weeks to return a diagnosis, a gap of greater than three months between birth date and diagnosis date was considered reasonable to presume a new diagnostic method had been added. A decision was made to not include these individuals in the average time to conversion calculation due to the uncertainty surrounding the timings of the diagnoses.

Overlap was expected between conversion categories, yet only one individual fit both a clinical and sweat chloride conversion. When considering the first sweat chloride result for the infants who converted to CF, the mean sweat chloride was 40mmol/L which is in the intermediate category. This is a higher average than the mean sweat chloride for the entire CFSPID population (29 mmol/L) which does suggest individuals with a higher sweat chloride are more likely to present with a clinical conversion.

This subgroup of individuals on average had more sweat tests performed than the CFSPID population. The presence of clinical symptoms may have made clinicians more vigilant in following sweat test guidelines.

This correlates with the short time between birth and diagnosis in these individuals and although they fit the CFSPID definitions it is likely that their clinical picture more represents a diagnosis of CF.

4.2.5.1 Persistent acute respiratory infection

The most common reason for clinical conversion was a persistent acute respiratory infection, which is a registry term. Individuals with CFSPID are known to have an increased prevalence of pseudomonas aeruginosa and staph aureus and increased pseudomonas infection is seen in individuals who convert to a diagnosis of CF (19). More information on the clinical presentation of these individuals is required to determine whether they were symptomatic or just had positive respiratory swabs.

Persistent acute respiratory infection was also seen alongside meconium ileus, and it is unclear how it fits with the clinical picture as meconium ileus usually occurs in the neonatal period and is indicative of CF not CFSPID2. This circumstance may have been considered CF from the neonatal period even though their genotype and sweat chloride concentration suggest CFSPID.

Furthermore, persistent acute respiratory infection was also seen in combination with electrolyte imbalance. Electrolyte disturbance often occurs at birth in infants with CF, for example, hyponatraemic dehydration and this could lead to a quick diagnosis of CF (2). The combination of clinical symptoms does suggest a diagnosis of CF even on a background of CFSPID.

The other methods of diagnosis, malnutrition and electrolyte imbalance are clinical symptoms that are often seen early in life whereas respiratory involvement is often seen later (7).

4.2.5.2 Malnutrition

Malnutrition was seen as a diagnosis method in the CFSPID population and can occur if cystic fibrosis is left untreated in the neonatal period (34). This also

suggests a diagnosis of CF regardless of the CFSPID genotype and sweat chloride concentration.

Overall, this is a difficult population to consider due to the lack of detail concerning the clinical picture of these patients and the lack of clarity regarding the diagnosis dates and if subsequent diagnosis methods have been added. More information is required on these individuals to make a definitive decision of a clinical conversion.

4.2.6 Reclassification of *CFTR* gene variants

The CFTR2 database was used to determine whether variants in our study population were reclassified, and no individuals were identified to have converted. Two variants were identified to have been reclassified however neither fit the criteria for a designation of CFSPID due to raised sweat chloride concentration.

4.2.7 Time to conversion

The median time to conversion was 39.8 months (3.3 years) which is very similar to other literature (18, 41, 64). The minimum time to conversion was the same day whereas the maximum time was just under 10 years. Whether the individual converted on their first repeat sweat test and the gap until the next sweat test is important to consider any delay in diagnosis. As mentioned above, only those with a sweat chloride are included in the calculation, not those with a clinical conversion.

When considering other studies, Ooi, Kharrazi and Terlizzi had a mean time of 1.8 years, 2.5 years, and 2 years respectively (18, 41, 64). Therefore, our time to conversion is slightly later than other studies and this may be due to a lack of a consistent application of a standard for repeat sweat testing in our population. A few individuals in our population had repeat sweat tests performed much later than guidelines indicate, and it is possible that if a sweat test was performed earlier, they may have been detected sooner.

There is a pattern of individuals converting late due to a delay in first repeat sweat test. Considering sweat test timing, these sweat tests may be performed due to a change in disease status for example clinical symptoms or may be a routine follow up sweat test. If sweat tests had been done more frequently this individual may have converted earlier with raised sweat chloride. They did have a repeat sweat test a month after their conversion which confirmed a sweat chloride in the range diagnostic for CF.

Some individuals had a sweat test performed yearly which results in much earlier conversions at three year which is more in accordance with the literature. Regular repeat sweat tests allows for conversions to be detected earlier which allows for the early initiation of treatment. The results and mean time to conversion show that children should be followed up at least once in the first two years and again at the age of three.

When considering the recently published guidelines, there seems to be a consensus that if an individual has a normal sweat chloride (<30) at age 6 then it is unlikely that they will convert later. This does rely on individuals having at least one sweat chloride before that age which only half of the CFSPID population in this study did.

That also raises the issue of what to do with CFSPID patients that have not converted by age six that have a normal or intermediate sweat chloride. As it is unlikely that they will convert, should they be discharged into the care of a primary care physician who can then monitor the patient for symptoms? The other option is to keep the children in the CF service until adolescence and monitor them periodically. Keeping these children in the CF service can place additional pressure on the service to provide appointments for potentially healthy children and regular appointments could lead to disruptions to schooling or their parent's work. Children who are deemed well are often lost to follow up as they or their parents remove themselves from the service.

One solution to this issue would be to only keep a record of those with an intermediate sweat chloride at age 6 and to discharge all those with only normal sweat tests. An intermediate sweat chloride is the only factor that has been identified that predicts conversion (7). When considering our study population, 80% of the individuals who converted had an intermediate sweat test 1.

4.3 Sweat Chloride Analysis

4.3.1 ECFS guidelines

The guidelines for the frequency of sweat testing are from the ECFS NSWG paper recently published providing a framework for the follow up of individuals and an outline of what investigations should take place at each age (13). Guidance on sweat tests states that sweat tests should be completed periodically throughout childhood and split into three time periods. These periods are in the first two years of life, preschool years (3-5 years) and at six years old (13).

In the first 2 years of life, a sweat test is indicated at 6 months and 2 years for those with an inconclusive diagnosis. Another sweat test can also be performed at 12 months if it is clinically indicated but the guidelines indicate that a sweat test at 2 years is the most discriminatory for a conversion to CF (13).

Considering our CFSPID population, many children have an initial diagnostic sweat test and then a further sweat test in the first year of life. There are few sweat tests at two years and even fewer sweat tests per individual in the preschool period with the majority not receiving a test during this time. The data on evaluation at six years old was analysed in a subset of the CFSPID population who had reached six years old and only a tenth received a sweat test at this age. The lack of sweat test data beyond the first year of life is concerning and suggests that clinicians are not aware of the guidelines which could lead to conversions being delayed or missed.

4.3.2 Missing sweat chloride data

There was a considerable amount of missing sweat chloride data in this population which suggests that many CFSPID patients are not followed up or sweat test data is not added to the registry. Clinicians may not add sweat data for CFSPID patients as they feel it does not belong in the CF registry. These individuals may not be removed from the registry, but clinicians do not add any follow-up information for the patients.

There were some irregularities with the sweat test data as certain individuals in the data set had values for sweat test two but missing values for sweat test one. The diagnostic sweat test data may be missing and then follow-up data was added in the correct field. It could also be that the diagnostic sweat test was added into the incorrect field.

On closer examination of the data on sweat tests, numerous children are entered on the registry without a sweat test performed. This is concerning as a diagnostic sweat should always be performed; it is possible that these patients had genotype analysis that showed two CF causing variants and therefore a diagnosis of CF was made without an accompanying sweat test or the sweat test data was not recorded on the registry.

Individuals diagnosed with CF are expected to have one sweat test performed at diagnosis. An additional sweat test may be done in preparation for *CFTR* modulator therapy to assess response to therapy. In contrast, individuals with CFSPID are expected to have multiple sweat tests periodically to consider conversion to a diagnosis of CF.

4.3.3 R117H/D1152H

It was deemed important to consider whether there are any patients with common VVCC that often cause CFSPID that either had no sweat test performed or no data on their sweat tests. This was to determine if there are any potential individuals

with CFSPID that we have not been able to identify due to a lack of sweat test data. The common VVCCs searched for were R117H and D1152H.

Considering those with no sweat test performed which consisted of 642 individuals, 21 had at least one copy of R117H and 5 had a least one copy of D1152H. This is concerning as these individuals should have had at least one sweat test to determine their disease status especially due to the uncertain nature of their *CFTR* variant. The outcome of these individuals is unclear, and it is possible that they were symptomatic at presentation or have another method of diagnosis. The lack of sweat tests also provides no information on follow-up and whether these individuals remain in the CF service.

When considering those with a sweat test performed but no data on the value, the two most common VVCCs were searched for again in the population. 24 individuals had at least one copy of R117H, and one had at least one copy of D1152H. These sweat tests may have had a normal or intermediate value which was not added to the registry, thus along with genotype information this individual has a designation CFSPID. If a child fits a designation of CFSPID, clinicians may recognise that these individuals should not be added to the registry and therefore no further information may be added. If this is the case, individuals with CFSPID have potentially been missed and not added to our population. Rather than leave an incomplete dataset on the registry, a solution would be to consider these individuals to have their diagnosis reversed therefore they are no longer considered to have CF.

4.3.4 Diagnostic sweat tests

There were a few individuals with multiple sweat tests on the same day. The Clinical and Laboratory Standards Institute (CLSI) guidelines recommend bilateral sequential sweat collection as this decreases the failure rate of the test and ensures a sufficient amount of sweat is collected for analysis (63). Bilateral testing also does not significantly increase the time the procedure takes nor increase discomfort

therefore may be favourable for a patient (63). However, it is considerably more expensive to perform bilateral testing.

In addition, performing both sweat tests on the same day does reduce the anxiety for the patient as repeat call-backs for sweat tests could leave patients with a sense that something is wrong, when an insufficient amount of sweat may have been collected during the first test. Steps should be taken to reduce anxiety for the patient wherever possible to speed up the diagnostic process, which can be a stressful time for families with children with suspected CF.

There are only a few individuals with two sweat tests recorded on the same day, therefore it is unclear if bilateral testing is routinely performed, but both sweat test results may not be added to the registry. An audit performed in a UK centre showed that only using sweat collected from a single arm did not miss any cases of CF compared to using both arms (63). Therefore, it is considered sufficient to only use one arm unless abnormalities in the testing are suspected (63).

Recording two different sweat tests on the same day can lead to diagnostic difficulties if there is a greater than 10mmol/L difference between the two measurements, which is not within the margin of error of the test. Intra-individual biological variation is also of importance when considering sweat test values, especially in tests with values close to cut-offs (63). The Guidelines for the Performance of the Sweat Test for the Investigation of Cystic Fibrosis in the UK also state that biological variation affects sweat test results more than imprecision in the test (22). Therefore, if two sweat tests had different classifications, for example, intermediate and raised it would be difficult to decide whether to identify this individual as CFSPID or CF. The infant would likely be treated as CF and the raised sweat test would be considered more important.

4.3.5 Normal vs intermediate sweat chloride concentration

The CFSPID population highlighted more individuals with a potential CFSPID designation with a normal sweat test than an intermediate sweat test. An increased

number of normal sweat tests being added to the registry was unexpected as it was thought that individuals closer to the cut-off for a diagnosis of CF would have been added as a clinician may believe that these individuals are more likely to convert.

4.3.6 Longitudinal sweat test data

Considering longitudinal sweat test data, the results of this study show that half of the individuals in the CFSPID population only have one sweat test recorded. This is disappointing as potential conversions may have been missed in this group due to a lack of follow-up. Furthermore, it meant that only half of the CFSPID population were able to be considered for a sweat chloride conversion. Clinicians may be assessing these patients based on their clinical picture and not through sweat testing which may explain the lack of longitudinal sweat test data.

In addition, when considering the timing of repeat sweat tests, 15% occur in the first 31 days from birth, which is not ideal when considering potential conversions as previous research suggests that conversions occur most frequently at around 2 years of age (18, 41, 64).

Therefore, it would be hoped that these individuals would have a further sweat test at around two years resulting in a total number of three sweat tests performed. Only 34% of the CFSPID population have a sweat test after these 31days, therefore it is unlikely.

Confirmatory sweat tests within 31 days are useful especially if the initial diagnostic sweat chloride is close to 60mmol/L and if it subsequently increases, early diagnosis can lead to the correct treatment being initiated sooner which results in better outcomes (35, 36).

There is some variability in the number of sweat tests performed in an individual with a potential diagnosis of CFSPID. This could be due to clinician preference and education on the CFSPID designation and expected follow-up.

Alternatively, these patients may have become symptomatic prompting a more thorough clinical investigation and guidelines suggest that investigation should be guided by the clinical picture of the patients(3, 55).

4.3.7 Average sweat chloride value

The mean sweat chloride value for sweat test one in the CFSPID population was 31mmol/L which is just above the cut-off for the normal sweat chloride concentration. It was expected that the average sweat chloride would be close to the cut off for CF (60mmol/L) as it seems more acceptable than those with borderline sweat chloride are placed on the registry. There are slightly more individuals in the normal group, but the mean sweat chloride is significantly skewed towards the normal population.

Sweat chloride values change with age as a study in individuals without CF showed their median sweat chloride increases over time until it plateaus at around age 20 (63). Therefore, this may suggest the need for 'age-related reference intervals' to make sure that only those with CF are picked up by sweat testing (63).

4.3.8 Age at first recorded sweat test

It was thought that the CFSPID population may have had their first sweat tests overwritten with later sweat tests; therefore, it was important to consider the age at first recorded sweat test to determine if the first sweat test added was likely to be the diagnostic test.

The mean age at the first recorded sweat test is 0.25 years (91 days) and a median of 0 years. The median suggests that a considerable number of sweat tests are performed at birth which is in line with the newborn screening programme. The newborn screening programme takes approximately six weeks (42 days) to confirm a diagnosis of CF and is elongated due to the second IRT measurement at 3 weeks (21 days). Therefore, it is possible that as the mean is significantly higher than this, some individuals have a first sweat test much later than is expected with the NBS

programme. This means that it is not possible to discount the fact that sweat tests could have been overwritten at a later date in a small number of individuals. The possibility of overwriting was raised by the registry team however no evidence of overwriting was found.

Considering the CFSPID population and the timings of the first sweat test, 173 (91%) had this test performed in the first year of life which is consistent with a diagnostic sweat test. When evaluating the second sweat test, 48 (49%) were done in the first year of life, this was increased to 71 (73%) by the second year of life, suggesting that a considerable number of second sweat tests were performed when the individual was less than two.

4.3.9 Time to next sweat test

Time to the next sweat test was important to consider this information could be used to highlight clusters of sweat tests and the times that individuals are more likely to be seen by clinicians.

To be in line with ECFS guidelines, a sweat test is recommended at 6 months and one year. In our study, the median time to the next sweat test is just over a year which is encouraging and suggests those who have multiple sweat tests may have them performed in line with guidelines. Only those with repeat sweat tests were able to be included in this population and there is a considerable population without a next sweat test who would not be meeting the ECFS guidelines.

It is concerning to consider children who had their next sweat test when they were older than two as there is a risk that they have already converted. Infrequent sweat testing can lead to delayed detection of conversions and the initiation of CF management.

4.3.10 Comparing sweat tests before and after the introduction of the CF screen positive, inconclusive diagnosis designation in 2014

The proportion of individuals in the CFSPID population with follow-up sweat tests was hoped to have improved following the introduction of the CFSPID designation in 2014.

The results comparing the two time periods before and after 2014 show almost identical trends in sweat tests and the proportion of the population with at least one sweat test. This demonstrated that sweat test practice has not been affected by the introduction of the CFSPID designation, which is interesting as CFSPID patients require significantly more sweat tests than the CF population. The new ECFS guidelines outlining the management of CFSPID patients provide more guidance for CF clinicians and highlight the importance of repeat sweat testing to identify conversions. Therefore, in the next couple of years, the proportion of individuals with sweat tests at each age will hopefully increase.

When breaking this population down further into each year from 2015-2019, there may have been a slight improvement in the proportion of sweat tests more recently in 2018 and 2019. It will be important to continue to monitor this trend over the next couple of years to see if there are further improvements that would indicate that the guidelines are being adhered to by clinicians.

4.3.11 Management of individuals with a designation of CF screen positive, inconclusive diagnosis

The next part of this study will investigate annual review data and see if these CFSPID patients are regularly followed up in the CF clinic. If it is found that the majority are regularly followed up in clinic for their annual reviews and their clinical picture assessed, it raises questions over why there is such a lack of follow-up sweat test data.

It also raises the question of who should oversee the care of these patients. It is the CF team that see these patients in clinic, which is appropriate for these patients due

to the risk of conversion, the need for sweat testing, knowledge of the CFTR variant databases and familiarity with CF symptoms. The CF clinic may not always the best place for these healthy children to have their follow-ups. The risk of converting from CFSPID to CF is relatively low (<10%) in this study and is similar in other studies, therefore do all children need this level of intensive follow-up when 90% will remain healthy and asymptomatic (18, 41, 64).

CF clinics are often in tertiary centres that are highly specialised to manage those with CF it is important that CFSPID patients do not overwhelm the resources of the clinics. As infants with CFSPID continue to be recognised, partnerships between primary and secondary care are important so that GPs can assist with the monitoring of individuals with CFSPID and be able to identify red flags that require a referral to the CF service.

If this is not feasible then GPs could help CF clinicians by keeping a record of CFSPID patients and reminding patients of their need to be routinely tested for conversion to CF. This could also be achieved with a CFSPID sub-registry which would also keep all the patients in one place, so they are not lost to follow-up.

There are also negative consequences of seeing healthy CFSPID infants in clinic, for instance, there is an increased risk of those with CFSPID testing positive for common CF pathogens, Staphylococcus aureus and Pseudomonas aeruginosa (11, 17, 18). It is possible that this could be at the same prevalence as is present in the normal population and would be detected on routine swabbing. It could also be due to multiple visits to CF clinics where the pathogens could be acquired.

In addition, there is the psychological effect of having to attend the CF clinic and the anxiety it may cause for patients and their families. A designation of CFSPID starts a painful waiting game for patients and families due to the ever-present risk of conversion in the first few years of life. This may replace the initial sense relief of parents may feel when they are told their child does not have CF.

4.3.12 Follow-up later in childhood

When considering the >6 years follow up, only 11% of the eligible population have a sweat test after six years of age. This is concerning as all individuals are supposed to have a clinical evaluation at 6 years including a sweat test, which is not occurring for much of the population as it is possible that the guidance is not recognised by many teams.

This population is the first cohort to be identified and managed as CFSPID as the designation was introduced in 2014 (5). Therefore, their management may not be in line with current guidance as clinicians may lack awareness of the follow-up required.

4.4 UK CF Registry

There is not yet international agreement on where to record individuals with CFSPID (7). Their presence on the CF registry is inappropriate and can overmedicalize these children. In addition, these individuals could be initiated on full CF management, which is not a good use of healthcare resources as these individuals are asymptomatic.

A dedicated CFSPID registry would be able to calculate the number of conversions, length of follow up and clinical outcomes such as pancreatic sufficiency and pseudomonas aeruginosa infection more accurately.

4.4.1 CF screen positive, inconclusive diagnosis field vs sub-registry

An adjustment to the CF registry that could demarcate CFSPID from CF would be the introduction of a field that classifies patients as either CF or CFSPID. This would be relatively easy to do as this study has already identified individuals that potentially have CFSPID. It would also make it easier to check if these patients are receiving adequate follow up in line with the European guidelines.

In addition, if a new CFSPID sub-registry is developed subsequently, these individuals can easily be transferred over without the need for extensive genotype evaluation of the registry having to be repeated.

One caveat is that these individuals only potentially have CFSPID and the labelling of them would require some clinical judgement. As new individuals are added to the registry, clinicians could have the option to add them with a label of CFSPID and this could lead to the identification of more individuals with CFSPID as clinicians would not be as apprehensive to add them to the registry due to fears they do not belong there.

A CFSPID registry provides a space separate from the CF registry where these patients can be recorded which could lead to greater accountability of clinicians to make sure that they routinely follow these patients up. A reminder system of the appointments suggested by the ECFS guidelines could lead to better outcomes for these patients when considering conversions.

A dedicated CFSPID registry would also assist research into these individuals and may help to better characterise variants that cause CFSPID and their prevalence. Furthermore, a significant amount of CFSPID data could help to predict those who may convert to a diagnosis of CF considering previous sweat test values, their genotype and other factors for example IRT levels as studied by Ooi (14).

4.4.2 Use of the UK CF registry

The UK CF Trust provide a report each year which provides headline statistics and figures on CF diagnosis, lung health, complications, and the use of different therapies (18). This report is important for patients, their families, and the wider healthcare team to highlight areas in which improvements are being made and areas that require more attention. In addition, it shows trends in treatment, what drugs are preferred and the performance of individual CF centres.

A similar report, although a substantial amount of work, has the potential to be important for CFSPID individuals. It would help clinicians understand how each centre deals with CFSPID patients and the number of individuals identified as CFSPID at each centre. In addition, it is important to consider how clinicians around the country treat those who potentially have CFSPID. It is possible that some centres give these individuals full CF treatment which is not advised and if this could be detected at an individual centre level then this could be prevented. Follow-up could be monitored and split at a centre level to highlight centres that are adhering to ECFS guidelines and those that may need to alter their approach to CFSPID patients.

4.5 Comparisons to other studies

The recent ECFS NSWG paper on guidance on the management of children with CRMS/CFSPID includes a table that outlines the key findings of similar recent studies (13). It summarises the results of prospective, retrospective and registry studies and it is useful to compare these results to what was found in this study.

4.5.1 Prospective Studies

First, two prospective studies performed by Ooi et al and Munck et al were considered: Ooi with a study into CRMS/CFSPID in Canada and Italy and Munck considering the population in France (41, 55).

4.5.1.1 Ooi

Ooi *et al* in Canada and Italy considered those who were born in 2007-2013 and had a median follow up of 2.2 years (41). This is a short follow up period as in our study the median time to conversion was just over 3 years. This may result in some conversions being missed in this study.

They reported their conversion rate to be 11% with a mean age of conversion of 1.8 years ± 1.2 years. This is significantly younger than our conversion age but when looking into the reason for conversion the majority were converting due to 2 CF-

causing variants, which could account for the younger conversion age as we did not find any individuals in our population that converted due to reclassified variants. These individuals also may have had more frequent sweat tests performed as this study was interested in conversions from CFSPID to a diagnosis of CF.

They reported the percentage of their population that has the genotype F508del/R117H to be 19.5% which is compared to 70% in our population. This could be due to greater ethnic diversity especially in the Italian population leading to a different prevalence of genotypes.

4.5.1.2 Munck

The second prospective study by Munck *et al* looked at those born from 2002-2009 in France and had a much longer mean follow up duration of 7.4 years (55). They reported a very high conversion rate of 44% with over half converting due to two CF causing variants. A larger number in this study converted due to increased sweat chloride compared to the previous prospective study, which is more in line with the results from this study.

This study had an F508del/R117H percentage of 43% which could suggest similarities in the ethnic diversity between the UK and French population.

4.5.2 Retrospective Studies

There are four retrospective studies included in the ECFS paper: Kharrazi *et al* with a study in California, Groves *et al* in Australia, Levy *et al* in Wisconsin, and Terlizzi *et al* in Tuscany (18, 56, 64, 65).

4.5.2.1 Kharrazi

The first retrospective study considered was Kharrazi who looked at those born between 2007 and 2012 in California and had a mean follow up period of 4.5 years (64). This is a reasonable follow-up period and if regular sweat tests are performed

to look for conversions it likely that the majority have been picked up by the end of this follow-up period.

They reported the percentage who converted to be 5.8% which is slightly lower than found in this study but a similar range unlike the Munck study mentioned previously (55). Most of their conversions were due to increased sweat chloride and they noted no conversions due to reclassified variants which is in accordance with our work. They noted a small group of conversion due to 'other criteria' which could have been the development of clinical symptom which was not present as a named option. Kharrazi had a mean conversion age of 2.5 years which is slightly lower than our study but similar.

4.5.2.2 Groves

The next study by Groves *et al* was a retrospective case-control study that followed those born in Australia between 1996 and 2010 for 10 years (56). This study recognised low numbers of individuals with CFSPID but with a high percentage conversion of 48%. The majority were due to "other criteria", it is not clear what these criteria are, but they may be a clinical diagnosis of CF.

They only identified a small number of individuals who converted due to an increased sweat chloride, but they did note that only 60% of this population had a repeat sweat test. This may sound like guidelines have not been followed as individuals are supposed to have routine sweat tests, but the median age of conversion is 0.19 years. Therefore, it is not indicated for these patients to have multiple sweat tests at this age.

In addition, it is noted that the definitions that were used to define CRMS/CFSPID are slightly different to those included in the ECFS paper which may explain the very young age of conversion (13). When considering the genotype for this population, 29% had a genotype of F508del/R117H. The Australian population may have more individuals that are not of Northern European descent and other variants could be the cause of the conversion.

4.5.2.3 Levy

Levy *et al* performed a retrospective cross-sectional study in Wisconsin, following up those born between 1994 and 2012 for eight years (65). This study had a high percentage of those with F508del/R117H with 63% having this genotype. No individuals were reported to have converted within this period.

It would have been expected that individuals would have converted by the end of the eight-year follow-up, but the CRMS/CFSPID population was small so perhaps a larger study size or a follow-up study would pick up conversions.

4.5.2.4 Terlizzi

Terlizzi et al performed a retrospective study in Tuscany, Italy considering those born between 2011 and 2016 (18). They followed this population up for a median time of 0.6 years. In this study, there were significantly more individuals identified as CFSPID than CF which could suggest that the newborn screening programme uses extended gene sequencing which picks up more CFSPID infants than using only biochemical screening.

Even though the median follow-up period is short, 10% of the individuals in this group converted to a diagnosis of CF which is similar to our study. All these individuals converted due to an increased sweat test which is in accordance with our work and the median time to conversion is 2 years with a range of 0.2 years to 4 years. Therefore, it seems that at least some of the population had longer follow-up.

There were 0 individuals in the CRMS/CFSPID population with the genotype F508del/R117H, which is not in agreement with our work. This further suggests that as a considerable amount of the Italian population may not be considered northern European, different variants are more prevalent. It would be interesting to consider the genotypes prevalent in this population as it is possible that R117H was not included in the gene panel.

4.5.3 CFF Registry Study

The study performed by Ren et al explored the Cystic Fibrosis Foundation (CFF) registry in the US (11). It considered those born in the US between 2010 and 2012 and included a follow-up period of a year. In this period, no conversions were noted which is interesting as this study had the largest CFSPID population out of all the studies recorded.

Considering the genotypes present in this population, 26% had the genotype F508del/R117H. This is less than expected for this population especially compared to the 70% noted in our study. In addition, the percentage was a lot higher in the Wisconsin study at 63% F508del/R117H, even though the studies were taking place simultaneously (65).

4.6 Newborn Screening Challenges

Newborn screening for CF in the UK faces a period of potential change, as momentum mounts to increase the number of variants identified through extended gene sequencing. The current UK protocol uses a small panel of variants compared to other countries but still retains good sensitivity (few missed cases). This may reflect the performance of the IRT assay when the DBS sample is collected a day 5, which is much later than other programmes. Also, despite the increasing diversity of the UK population the prevalence of F508del remains high and most infants with CF have at least one variant identified on the limited first panel. For those with rarer variants, the safety arm of the protocol enables identification of infants with rare variants. At the moment, the UK programme has both the best sensitivity and positive predictive value of most programmes in Europe (38). The argument for moving to a protocol that uses extended *CFTR* gene relates to the aim of improving timelines (through only having a requirement for one sample) and possibly improving specificity, especially if the programme moved to not reporting carriers and infants with two variants when one of the variants is a VVCC. Obviously, this raises issues of non-disclosure of results and a large stakeholder

engagement programme is underway in the UK to evaluate public opinion on this proposal.

If next generation sequencing is employed to provide extended *CFTR* gene analysis as part of the protocol, this approach may result in the recognition of many more (possibly ten-fold) infants with a CFSPID designation. An approach that may reduce this impact is described below.

One way to reduce the numbers of carriers and potential CFSPID designations is to not report them or inform families of these results. For carriers, this could be used if one *CFTR* variant has been identified but EGS has identified no further variants. This result would then not be reported to families with the aim to not overmedicalise healthy individuals. This does have ethical consequences as individuals and their families may want to be aware of their carrier status as it may influence reproductive decision.

Similarly, variants of varying clinical consequence (VVCCs) could also not be reported which has repercussions for the CFSPID population as the majority have at least one VVCC. Not reporting VVCCs will result in fewer individuals been identified with potential CFSPID, which is concerning due to the risk of conversions. Individuals with CFSPID will not be followed up in line with ECFS guidelines which could result in the delayed detection of conversion.

To conclude, there are a wide variety of approaches to NBD for CF across the globe and, to some degree, this reflects a countries geographical and social features, as well as existing healthcare infrastructure. Countries have to make decisions when designing their newborn screening protocols to decide whether they want to prioritise detection of CF or minimise CFSPID numbers. As more countries move to extended gene sequencing, the number of individuals with a designation of CFSPID is likely to significantly increase which provides challenges for CF centres in terms of the logistics of follow-up and the question of whether to report all these individuals.

4.7 Strengths and Limitations

4.7.1 Strengths

This study included all new diagnoses of CF by newborn screening within the specified period which were able to be screened to determine whether they potentially have CFSPID. This includes data from all the CF centres in the UK and the registry is estimated to cover 99% of the UK CF population, which results in a large sample size and a study population that is representative of the UK.

This study was largely helped by the publication of the ECFS guidelines on the management of children with CFSPID (13). This document guided the definitions of CFSPID, comparison of studies looking into CFSPID and detailed guidance on when to follow-up these individuals and what investigations should be done during these clinical encounters. It also provided a benchmark for the number of follow up appointments that were compared to our population. The clear criteria for identifying individuals with CFSPID helped to determine whether these individuals should be present on the registry and assisted with the detection of those with potential CFSPID.

4.7.2 Limitations

This study was only able to recognise those with CFSPID who had been incorrectly added to the CF registry. Therefore, it is likely that the CFSPID population has been underestimated as individuals with CFSPID may not have been added to the CF registry. A way to capture the full CFSPID population in the UK would be to contact CF centres and use patient notes to identify individuals who have not been added to the registry. this would be particularly important to provide complete information for the new CFSPID sub-registry.

The registry contains a considerable amount of missing data as information has not been added in full by CF centres. Therefore, this could have led to an underestimation of the number of individuals with a potential designation of

CFSPID and the number of conversions noted. Both these points may have resulted in a considerable number of individuals with CFSPID not included in our study which may affect the conclusions regarding conversions and sweat test analysis. A follow-up to this study in 10 years would be interesting to consider improvements in data collection, to observe changes in the number of individuals with a designation of CFSPID and whether these individuals receive follow up in line with ECFS guidelines.

R117H poly-T data has only been recorded by the UK CF registry from 2016, therefore for the years prior to this, the variant classification of the R117H haplotype is unknown. This could mean that individuals with CF have been included in this CFSPID population leading to an overestimation of the group. In addition, as the variant R117H is the most prevalent in the UK population this could affect a large proportion of the population.

The CFSPID population is only an estimation and does not include CFSPID individuals who are not present on the UK CF registry, therefore the calculations made regarding length of follow-up and the proportion of the population with at least one sweat test may not be representative of the entire population. It is possible that the population not added to the registry may have had more sweat tests performed in line with the ECFS guidance. For this population to have their sweat test data considered entirely, CF centres would need to be contacted and data for all individuals requested.

4.7.3 Future work

This study sets the groundwork for future research into the UK CFSPID population and is one phase of a larger study. The next phase of this study will delve into clinical parameters and compare the trajectories of CFSPID patients compared to those with CF.

One parameter that will be investigated is pseudomonas aeruginosa infection. The literature suggests that the rate of pseudomonas infection is increased in CFSPID

patients compared to the normal population and in addition, infection is found more often in those individuals who convert to a diagnosis of CF.

In addition, pancreatic sufficiency will be investigated as other studies have found that the CFSPID population are invariably pancreatic sufficient whereas those with CF are usually pancreatic insufficient. It will be of interest to consider those individuals who have converted from CFSPID to CF to see whether this change in disease status has affected pancreatic sufficiency. As individuals who have converted often have a milder phenotype, pancreatic status may not change.

Similarly, weight trajectory will also be analysed to compare the CFSPID and CF population. It would also be interesting to also compare the CFSPID and CF population alongside the normal population to determine whether the *CFTR* variants, regardless of the phenotype they cause affect weight trajectory.

Lung function, for example, FEV1, will be compared to that of CF children to study whether lung function is affected in those with CFSPID and whether it shows any decline. The effect of CFSPID on lung function does not seem to have been mentioned extensively in the literature on CFSPID.

Finally, the number of hospital admissions will be studied as admissions would not be routinely expected in the CFSPID population as they are s healthy asymptomatic population. It may be possible that those who have a hospital admission may be more likely to convert to a diagnosis of CF therefore it may be possible to use it as a screening tool.

5. Conclusion

To conclude, there are 185 individuals included on the UK CF Registry without a clear diagnosis of CF. This is a considerable number which is inappropriate and may lead to the overtreatment and medicalisation of healthy children. It seems that the UK programme recognises a smaller proportion of infants with CFSPID compared to other national programmes especially those using extended gene sequencing.

The number of individuals with a potential CFSPID designation added to the registry since the publication of the designation has increased. The first part of the study (2008-2014) has a large amount of missing data which may account for the reduced number of individuals with CFSPID being identified. The recording of sweat test data is significantly worse than the recording of genotype information with a considerable group of individuals without a diagnostic sweat test added to the registry.

Furthermore, sweat test data suggest several of these CFSPID patients are not followed up according to the ECFS guidance. This has highlighted the need for CF centres to improve their follow up of these individuals to make sure that conversions to CF are not missed and are detected as early as possible. The CFSPID designation was only introduced in 2014 and education is required for clinicians in terms of how to diagnose, follow up and manage those with this condition. This can be done by further promoting the ECFS guidelines to UK CF centres so that clinicians have an increased awareness of the management and follow-up as recommended by the ECFS. The guidelines also provide detailed information regarding criteria to qualify as a designation of CF which may not have been available early on in this study.

Conversions present a unique challenge to clinicians as they represent a change in disease status. The recent ECFS guidelines provide routine points for follow-up and set out the investigations expected at each age, and it is hoped that a standardised

follow-up regime will result in the earlier detection of conversions to CF allowing for earlier initiation of appropriate management.

The best place to record individuals with CFSPID is a designated sub-registry that could make up part of the UK CF registry. The numbers of individuals diagnosed with CFSPID is unlikely to decrease unless there are changes to the UK CF screening protocol therefore it is essential that these patients are recorded adequately and receive appropriate follow-up.

One way to ensure appropriate management of this group is to introduce specialised pathways of care for CFSPID individuals. These pathways would not include CF therapies and treatments but rather the follow-up timeline and access to other services such as psychology. It is important to consider the psychological effect this diagnosis may have on these individuals. The anticipation of conversion to CF can be a burden on patients and their families and they should be reassured that the clinical phenotype after conversion is a milder atypical form of CF.

This is a challenging situation and the initial cohort identified with CFSPID have not yet reached adolescence. With about 10% of the population converting to a diagnosis of CF it is possible that this population is being overmedicalised and it is possible that individuals on the registry with potential CFSPID may be receiving full CF treatment. Treatment with multiple unnecessary medications could be harmful to children and could lead to a range of side effects.

It is hoped that in the future, there will be ways to predict conversions whether that be biomarkers such as IRT, specific genotypes or a pattern of sweat chloride results. This will only become clear with increased studying and monitoring of the CFSPID population which makes recording the individuals in a place easily accessible to researchers of paramount importance. If conversions can be predicted, then the other 90% of the CFSPID population would not have to undergo regular sweat testing and clinical evaluation and would not incur the psychological stress of awaiting a disease that might not occur.

Conversions that have already been detected seem to occur in the first few years of life. If this could be definitively confirmed, this allows clinicians to discharge individuals when they pass this threshold to avoid overmedicalising children for longer than is necessary. There is some agreement between studies so far regarding time to conversion but until sweat testing improves, this time may be overestimated.

This all points to further research into this population and the improvement of current management practices. As there is now much more guidance and information on this condition than there was in the earlier years of this study it is hoped that this will result in better outcomes regarding recording and follow-up.

In conclusion, we have recognised a relatively large population of children on the registry with no clear CF diagnosis, who have a potential designation of CFSPID. Only a small proportion convert to a CF diagnosis. This highlights the urgent need for a CFSPID registry or sub-registry and better management consistent with ECFS guidance.

References

1. UK Newborn Screening Programme Centre. A laboratory guide to newborn screening in the UK for cystic fibrosis, 4th edn 2014 [Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/397726/Cystic_Fibrosis_Lab_Guide_February_2014_v1.0_12_.pdf.
2. Farrell PM, White TB, Ren CL, Hempstead SE, Accurso F, Derichs N, et al. Diagnosis of Cystic Fibrosis: Consensus Guidelines from the Cystic Fibrosis Foundation. J Pediatr. 2017;181s:S4-S15.e1.
3. Barben J, Southern KW. Cystic fibrosis screen positive, inconclusive diagnosis. Curr Opin Pulm Med. 2016;22(6):617-22.
4. Rock MJ. How to define CRMS/CFSPID conversion to CF. Pediatr Pulmonol. 2020;55(7):1548-9.
5. Munck A, Mayell SJ, Winters V, Shawcross A, Derichs N, Parad R, et al. Cystic Fibrosis Screen Positive, Inconclusive Diagnosis (CFSPID): A new designation and management recommendations for infants with an inconclusive diagnosis following newborn screening. J Cyst Fibros. 2015;14(6):706-13.
6. Munck A, Delmas D, Audrézet MP, Lemonnier L, Cheillan D, Roussey M. Optimization of the French cystic fibrosis newborn screening programme by a centralized tracking process. J Med Screen. 2018;25(1):6-12.
7. Bell SC, Mall MA, Gutierrez H, Macek M, Madge S, Davies JC, et al. The future of cystic fibrosis care: a global perspective. Lancet Respir Med. 2020;8(1):65-124.
8. Taylor-Robinson D, Archangelidi O, Carr SB, Cosgriff R, Gunn E, Keogh RH, et al. Data Resource Profile: The UK Cystic Fibrosis Registry. International Journal of Epidemiology. 2018;47(1):9-10e.
9. Naehrlich L. The Changing Face of Cystic Fibrosis and Its Implications for Screening. Int J Neonatal Screen. 2020;6(3):54.
10. Schlüter DK, Southern KW, Dryden C, Diggle P, Taylor-Robinson D. Impact of newborn screening on outcomes and social inequalities in cystic fibrosis: a UK CF registry-based study. Thorax. 2020;75(2):123-31.
11. Ren CL, Fink AK, Petren K, Borowitz DS, McColley SA, Sanders DB, et al. Outcomes of infants with indeterminate diagnosis detected by cystic fibrosis newborn screening. Pediatrics. 2015;135(6):e1386-92.
12. Southern KW, Mérelle MM, Dankert-Roelse JE, Nagelkerke AD. Newborn screening for cystic fibrosis. Cochrane Database Syst Rev. 2009;2009(1):Cd001402.
13. Barben J, Castellani C, Munck A, Barry L, Davies J, de Winter K, et al. Updated guidance on the management of children with Cystic Fibrosis Transmembrane Conductance Regulator-Related Metabolic Syndrome/Cystic Fibrosis Screen Positive, Inconclusive Diagnosis (CRMS/CFSPID) For the European CF Society Newborn Screening Working Group (ECFS NSWG) J Cyst Fibros. 2020;16(56).
14. Ooi CY, Sutherland R, Castellani C, Keenan K, Boland M, Reisman J, et al. Immunoreactive trypsinogen levels in newborn screened infants with an inconclusive diagnosis of cystic fibrosis. BMC Pediatrics. 2019;19(1):369.
15. Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. Nat Rev Genet. 2015;16(1):45-56.
16. Fedick AM, Zhang J, Edelmann L, Kornreich R. Prenatal Diagnosis of Cystic Fibrosis. Methods Mol Biol. 2019;1885:221-31.

17. Castellani C, Massie J, Sontag M, Southern KW. Newborn screening for cystic fibrosis. Lancet Respir Med. 2016;4(8):653-61.
18. Terlizzi V, Mergni G, Buzzetti R, Centrone C, Zavataro L, Braggion C. Cystic fibrosis screen positive inconclusive diagnosis (CFSPID): Experience in Tuscany, Italy. J Cyst Fibros. 2019;18(4):484-90.
19. Munck A. Inconclusive Diagnosis after Newborn Screening for Cystic Fibrosis. Int J Neonatal Screen. 2020;6(1):19.
20. Castellani C, Assael BM. Cystic fibrosis: a clinical view. Cell Mol Life Sci. 2017;74(1):129-40.
21. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science. 1989;245(4922):1066-73.
22. Nishida K, Smith Z, Rana D, Palmer J, Gallicano GI. Cystic fibrosis: a look into the future of prenatal screening and therapy. Birth Defects Res C Embryo Today. 2015;105(1):73-80.
23. Scobie G, Woodroffe B, Fishel S, Kalsheker N. Identification of the five most common cystic fibrosis mutations in single cells using a rapid and specific differential amplification system. Mol Hum Reprod. 1996;2(3):203-7.
24. De Boeck K, Zolin A, Cuppens H, Olesen HV, Viviani L. The relative frequency of CFTR mutation classes in European patients with cystic fibrosis. J Cyst Fibros. 2014;13(4):403-9.
25. Leung DH, Heltshe SL, Borowitz D, Gelfond D, Kloster M, Heubi JE, et al. Effects of Diagnosis by Newborn Screening for Cystic Fibrosis on Weight and Length in the First Year of Life. JAMA Pediatr. 2017;171(6):546-54.
26. The Clinical and Functional TRanslation of CFTR (CFTR2). 2011 [Available from: http://cftr2.org.
27. Claustres M, Thèze C, des Georges M, Baux D, Girodon E, Bienvenu T, et al. CFTR-France, a national relational patient database for sharing genetic and phenotypic data associated with rare CFTR variants. Hum Mutat. 2017;38(10):1297-315.
28. Castellani C. CFTR2: How will it help care? Paediatr Respir Rev. 2013;14 Suppl 1:2-5.
29. Vertex Pharmaceuticals. CF Source [Available from: https://www.cfsource.co.uk.
30. Smyth AR, Bell SC, Bojcin S, Bryon M, Duff A, Flume P, et al. European Cystic Fibrosis Society Standards of Care: Best Practice guidelines. Journal of Cystic Fibrosis. 2014;13:S23-S42.
31. Crossley JR, Elliott RB, Smith PA. Dried blood spot screening for cystic fibrosis in the newborn. The Lancet. 1979;1(8114):472-4.
32. Andermann A, Blancquaert I, Beauchamp S, Déry V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. Bulletin of the World Health Organization. 2008;86(4):317-9.
33. Course CW, Hanks R. Newborn screening for cystic fibrosis: Is there benefit for everyone? Paediatr Respir Rev. 2019;31:3-5.
34. Balfour-Lynn IM. Newborn screening for cystic fibrosis: evidence for benefit. Arch Dis Child. 2008;93(1):7-10.

35. Barben J, Southern KW. Why Do We Screen Newborn Infants for Cystic Fibrosis. Int J Neonatal Screen. 2020;6(3):1-5.
36. Scotet V, Gutierrrez H, Farrell PM. Newborn Screening for CF across the Globe—Where Is It Worthwhile? Int J Neonatal Screen. 2020;6(1).
37. Sims EJ, Mugford M, Clark A, Aitken D, McCormick J, Mehta G, et al. Economic implications of newborn screening for cystic fibrosis: a cost of illness retrospective cohort study. Lancet. 2007;369(9568):1187-95.
38. Barben J, Castellani C, Dankert-Roelse J, Gartner S, Kashirskaya N, Linnane B, et al. The expansion and performance of national newborn screening programmes for cystic fibrosis in Europe. J Cyst Fibros. 2017;16(2):207-13.
39. Barben J, Chudleigh J. Processing Newborn Bloodspot Screening Results for CF. Int J Neonatal Screen. 2020;6(2):25.
40. Bergougnoux A, Lopez M, Girodon E. The Role of Extended CFTR Gene Sequencing in Newborn Screening for Cystic Fibrosis. Int J Neonatal Screen. 2020;6(1).
41. Ooi CY, Castellani C, Keenan K, Avolio J, Volpi S, Boland M, et al. Inconclusive diagnosis of cystic fibrosis after newborn screening. Pediatrics. 2015;135(6):e1377-85.
42. Rock MJ, Mischler EH, Farrell PM, Lee-Jen W, Bruns T, Hassemer DJ, et al. Newborn screening for cystic fibrosis is complicated by age-related decline in immunoreactive trypsinogen levels. Pediatrics. 1990;85:1001-7.
43. Edmondson C, Grime C, Prasad A, Cowlard J, Nwokoro CEC, Ruiz G, et al. Cystic fibrosis newborn screening: outcome of infants with normal sweat tests. Arch Dis Child. 2018;103(8):753-6.
44. Castellani C. Newborn Screening for Cystic Fibrosis: Over the Hump, Still Need to Fine-Tune it. Int J Neonatal Screen. 2020;6(3):5-7.
45. Foil KE, Powers A, Raraigh KS, Wallis K, Southern KW, Salinas D. The increasing challenge of genetic counseling for cystic fibrosis. J Cyst Fibros. 2019;18(2):167-74.
46. Wilfond B, Rothenberg LS. Ethical issues in cystic fibrosis newborn screening: from data to public health policy. Curr Opin Pulm Med. 2002;8:292-6.
47. Scotet V, de Braekeleer M, Roussey M, Rault G, Parent P, Dagorne M, et al. Neonatal screening for cystic fibrosis in Brittany, France: assessment of 10 years' experience and impact on prenatal diagnosis. Lancet. 2000;356(9232):789-94.
48. Sommerburg O, Hammermann J. Pancreatitis-Associated Protein in Neonatal Screening for Cystic Fibrosis: Strengths and Weaknesses. Int J Neonatal Screen. 2020;6(2):28.
49. Sarles J, Giorgi R, Berthézène P, Munck A, Cheillan D, Dagorn JC, et al. Neonatal screening for cystic fibrosis: comparing the performances of IRT/DNA and IRT/PAP. J Cyst Fibros. 2014;13(4):384-90.
50. Krulišová V, Balaščaková M, Skalická V, Piskáčková T, Holubová A, Paděrová J, et al. Prospective and parallel assessments of cystic fibrosis newborn screening protocols in the Czech Republic: IRT/DNA/IRT versus IRT/PAP and IRT/PAP/DNA. Eur J Pediatr. 2012;171(8):1223-9.
51. Southern KW. Determining the optimal newborn screening protocol for cystic fibrosis. Thorax. 2012;67(4):281.

52. Rowe SM, Clancy JP, Wilschanski M. Nasal potential difference measurements to assess CFTR ion channel activity. Methods Mol Biol. 2011;741:69-86.
53. Derichs N, Sanz J, Von Kanel T, Stolpe C, Zapf A, Tümmler B, et al. Intestinal current measurement for diagnostic classification of patients with questionable cystic fibrosis: validation and reference data. Thorax. 2010;65(7):594.
54. de Poel E, Lefferts JW, Beekman JM. Intestinal organoids for Cystic Fibrosis research. J Cyst Fibros. 2020;19 Suppl 1:S60-s4.
55. Munck A, Bourmaud A, Bellon G, Picq P, Farrell PM. Phenotype of children with inconclusive cystic fibrosis diagnosis after newborn screening. Pediatr Pulmonol. 2020;55(4):918-28.
56. Groves T, Robinson P, Wiley V, Fitzgerald DA. Long-term outcomes of children with intermediate sweat chloride values in infancy. J Pediatr. 2015;166(6):1469-74.e1-3.
57. Mérelle ME, Huisman J, Alderden-van der Vecht A, Taat F, Bezemer D, Griffioen RW, et al. Early versus late diagnosis: psychological impact on parents of children with cystic fibrosis. Pediatrics. 2003;111(2):346-50.
58. Seror V, Cao C, Roussey M, Giorgi R. PAP assays in newborn screening for cystic fibrosis: a population-based cost-effectiveness study. J Med Screen. 2016;23(2):62-9.
59. Armstrong RE, Frith L, Ulph FM, Southern KW. Constructing a Bioethical Framework to Evaluate and Optimise Newborn Bloodspot Screening for Cystic Fibrosis. Int J Neonatal Screen. 2020;6(2).
60. The Clinical and Functional TRanslation of CFTR (CFTR1). 1989 [Available from: http://genet.sickkids.on.ca.
61. National Guideline A. National Institute for Health and Care Excellence: Clinical Guidelines. Cystic Fibrosis: Diagnosis and management. London: National Institute for Health and Care Excellence (UK)
Copyright © NICE 2017.; 2017.
62. Carr Sn, Cosgriff R, Charman S, Lee A, McClenaghan E. UK Cystic Fibrosis Registry Annual Data Report 2019. 2019.
63. Guidelines for the Performance of the Sweat Test for the Investigation of Cystic Fibrosis in the UK. 2014 [Available from: http://www.exeterlaboratory.com/images/sweat-guideline-v2-1.pdf.
64. Kharrazi M, Yang J, Bishop T, Lessing S, Young S, Graham S, et al. Newborn Screening for Cystic Fibrosis in California. Pediatrics. 2015;136(6):1062-72.
65. Levy H, Nugent M, Schneck K, Stachiw-Hietpas D, Laxova A, Lakser O, et al. Refining the continuum of CFTR-associated disorders in the era of newborn screening. Clin Genet. 2016;89(5):539-49.
66. Giordon E, Costes B, Cazeneuve C, Ghanem N, Goossens M. CFTR1.1995. [Available from: http://www.genet.sickkids.on.ca/MutationDetailPage.external?sp=406.
67. Onay T, Kirdar B, Zielenski J, Markiewicz D, Tsui L-C. CFTR11996 [Available from: http://www.genet.sickkids.on.ca/MutationDetailPage.external?sp=583.
68. Jézéquel P. CFTR11995 [Available from: http://www.genet.sickkids.on.ca/MutationDetailPage.external?sp=206.

69. Khajuria R, Kabra M, Garg P, Shastri S. CFTR12007 [Available from: http://www.genet.sickkids.on.ca/MutationDetailPage.external?sp=1644.
70. Seia M, Padoan R. CFTR12002 [Available from: http://www.genet.sickkids.on.ca/MutationDetailPage.external?sp=1294.

Appendices

Data request form

UK CF Registry
Data Request Form

UK CF Registry data request form

All requests for data or information from the UK CF Registry must be submitted on this form to registry@cysticfibrosis.org.uk .

Before starting a request please make sure you have read the UK CF Registry Data Request Policy, available at https://www.cysticfibrosis.org.uk/the-work-we-do/uk-cf-registry/apply-for-data-from-the-uk-cf-registry.

Please note that researchers will need to secure local R&D approval before submission – this will be mandatory from January 2020.

Research team

Provide details of everyone in the research team requesting data. Add more rows to the table if required.

Full name	Job title	Institution	Email address
Kevin Southern	Physician	University of Liverpool	kwsouth@liv.ac.uk
Siobhan Carr	Physician	RLCH, Brompton	s.carr@rbht.nhs.uk
Daniela Schlueter	Lecturer	University of Liverpool	d.k.schlueter@liverpool.ac.uk
Ellie Russo	MPhil Student	University of Liverpool	e.russo@liverpool.ac.uk
Jane Chudleigh	Reader	University of London	J.Chudleigh@city.ac.uk

Title of project / research question

How many children on the UK CF registry have an unclear diagnosis of CF following a positive newborn bloodspot screening (NBS) result and what was the impact of publication of the CFSPID designation in 2014?

Plain English summary (100 words max)
If your request is granted this and the title of your project will be published on https://www.cysticfibrosis.org.uk/the-work-we-do/uk-cf-registry/apply-for-data-from-the-uk-cf-registry

Infants with an unclear diagnosis after a positive newborn screening result for CF have the designation CF Screen Positive, Inconclusive Diagnosis (CFSPID). It is possible that CFSPID infants are being included in the UK CF Registry despite not fulfilling the criteria for a CF diagnosis. In this project we will explore the UK CF Registry to determine the extent of this problem, monitor for change over time and prepare for the new CFSPID registry (being constructed by the UK CF Trust) and in response to the new European guidelines on the management of infants with CFSPID.

Description and rationale of project (300 words max)

The national NBS programme for CF was implemented in 2007. We will assess if infants with an unclear diagnosis after a positive NBS result have been entered onto the registry and monitor their progress.
There are three phases to this project

Phase 1; will examine the number of infants with an unclear diagnosis of CF that are recorded on the Registry after a positive NBS result. We will record this year on year from 2007 and assess the extent of the problem and the impact of publication of the CFSPID designation in 2014. We will also interrogate the data to assess the progress of these infants, with respect to diagnostic evolution (do they convert to a clear CF diagnosis?).
Phase 2; Dependent on the results of phase 1 we will compare the clinical progress of infants identified with an unclear diagnosis compared to those with a clear diagnosis of CF (comparing weight trajectory, acquisition of *Pseudomonas aeruginosa* and early respiratory function measures (ppFEV1)).
Phase 3; We will triangulate aggregated data from the Registry with the recording of a CFSPID designation on the national NBS database collected by PHE. These data are anonymized, but we will explore methods to link with data recorded on the Registry to provide a clearer picture of the processing of these infants after their positive NBS result. These data will enable us to inform practice across the UK in preparation for the new CFSPID sub-Registry. In addition, they will provide valuable information to assess the extent of the problem with respect to the management of these infants, some of whom are now approaching adult life. This is particularly timely as new ECFS guidance on the management of children with CFSPID is about to be published.
We are proposing a staged approach to efficiently use the registry data.

Intended use(s) of the data
Including anticipated publications, presentations or reports

- To determine the number of infants with an unclear diagnosis on the UK CF Registry
- To compare this to national screening data
- To evaluate diagnostic evolution
- To compare clinical progress
- To inform national practice and improve the management and lives of these families
- To inform the new CFSPID sub-Registry

Name and qualifications of person performing the statistical analysis

Ellie Russo will be undertaking the principal analysis of the dataset with support from Dr Daniela Schlüter, who has published important work from the registry showing the benefit of NBS for CF, but the lack of impact on health inequalities (PMID, 31771956)

Please name any funding bodies that are supporting this work. This should include names of pharmaceutical companies where relevant

Not applicable

X	**Check this box if no funding has been received to support this work**

Data set type required (TICK ONE)	
X	**Check here to request data from the UK CF Registry dataset**
	Check here to request data cleaned by the EpiNet Strategic Research Centre (available up to 2017)

Exact details of data required	
Year(s) required:	2007-2019
Cohort required:	All new diagnosis
Variables required:	For each patient

	Baseline data - DOB - Gender - Ethnicity - Birth weight - Date of diagnosis - NBS result (all fields) - Screening lab (country) - Sweat test results and date (all results) - Genotype - Fecal elastase result - Social deprivation score Annual review data - Number of annual reviews - Pancreatic status - Weight - FEV1 - Age at acquisition of Pseudomonas aeruginosa - Number of hospital admissions
Project deadline:	As soon as possible please

X	Check this box if you wish to receive communications via email from the Cystic Fibrosis Trust

If your data request is approved we would be grateful if you could send the registry copies of any papers or abstracts where this data has been used

Data amendment form

UK CF Registry
Data Request Amendment Form

UK CF Registry data request amendment form
All amendments to original requests for data or information from the UK CF Registry must be submitted on this form to registry@cysticfibrosis.org.uk . Data Sharing Policy: https://www.cysticfibrosis.org.uk/the-work-we-do/uk-cf-registry/apply-for-data-from-the-uk-cf-registry

Lead Applicant

Full name	**Job title**	**Institution**	**Email address**
Kevin Southern	Physician	University of Liverpool	kwsouth@liv.ac.uk
Daniela Schlueter	Lecturer	University of Liverpool	d.k.schlueter@liverpool.ac.uk
Original Data Request Reference no:		**Date of Amendment Request:**	
404		09/02/2021	
Full Description of Amendment / Additional Researcher			
The initial analysis highlighted the need for some further variables we would like to request.			
Additional Researcher	**Job title**	**Institution**	**Email address**

Any additional variables not included on original request
- Diagnosis Method (all fields) - GenotypeQty - Genotype PolyT Tract variables (5T,7T, 9T and NK for Mut1 and Mut2)

Exact details of additional data required	
Year(s) required:	2007-2019
Cohort required:	All new diagnosis
Variables required:	- Diagnosis Methods (all fields) - GenotypeQty - Genotype PolyT Tract variables (5T,7T, 9T and NK for Mut1 and Mut2)
Project deadline:	As soon as possible please

Full methods of diagnosis

Below is the full list of diagnosis methods that are able available on the UK CF registry. From the full list, those that were considered clinical method were selected and are highlighted in the clinical methods of diagnosis list.

Full methods of diagnosis list:

- Bronchiectasis
- Electrolyte imbalance
- Family history
- Fertility
- Genotype
- Liver disease
- Malnutrition
- Meconium ileus
- Nasal polyps
- Newborn screening
- Oedema
- Other
- Pancreatitis
- Persistent acute respiratory infection
- Prenatal
- Rectal prolapse
- Steatorrhea

Clinical methods of diagnosis list:

- Persistent acute respiratory infection
- Oedema
- Electrolyte imbalance
- Malnutrition
- Bronchiectasis
- Pancreatitis

- Fertility
- Liver disease
- Nasal polyps
- Rectal prolapse
- Steatorrhea

Statistical Analysis Plan

Objective 1: Identifying the number of children with CF screen positive, inconclusive diagnosis on the UK CF Registry

- Study Design:
    - Study Type: retrospective cohort study
    - Data Source: UK CF Registry data
    - Study Population (eligibility criteria): all children in the CF Registry born in/after 2008 with a positive newborn screening result
- Outcome:
    - *Outcome*: unclear diagnosis of CF (CF screen positive, inconclusive diagnosis - CFSPID)
    - *Definition of outcome*:
        - Normal sweat chloride (<30mmol/L) + 2 CFTR variants (at least one with unclear phenotype)
        - Intermediate sweat chloride (30-59mmol/L) + one/no CFTR variants
        - Diagnosis reversed due to CFSPID
- Variables used to determine this in the UK CF Registry:
    - S03sweattestperformed
    - S03sweattestvalue01 – 05
    - S03sweattestdate01 – 05
    - S02genotypemut1
    - S02genotypemut1specify
    - S02genotypemut2
    - S02genotypemut2specify

- S03diagnosisreversed
- S03DiagnosisReversedDate
- S03DiagnosisReversedReason
- S03DiagnosisReversedReasonOtherDetails
- Other characteristics of interest: Birth_year

- Methods:
  - To identify individuals with CFSPID, the study population will be filtered into normal sweat chloride, intermediate sweat chloride and raised sweat chloride using the variables s03sweattestvalue01 and s03sweattestvalue02 and s03sweattestvalue03 (if within 31 days of the first sweat test). The difference between only using sweat test one and all sweat tests within 31 days will be highlighted in a table.
  - The individuals in the normal and intermediate categories will have their genotype screened using S02genotypemut1, S02genotypemut1specify, S02genotypemut2, S02genotypemut2specify
  - Information regarding genotypes will be sourced from CFTR2 database and definition of CFSPID from the recent ECFS NSWG paper.
  - Diagnosis reversal will be considered using the variables noted above and if any reversal reason was CFSPID they will be added to the CFSPID population.
  - For every year between 2007-2019 we will estimate the total number of children diagnosed by NBS and the number and the proportion of children that have an unclear diagnosis as identified by the procedure above. These results will be reported in a table and bar chart.
  - The population will be split pre- and post-2014 to determine whether number of CFSPID has changed since the designation was introduced
  - The genotypes of the individuals in the CFSPID population will be detailed to consider trends in the variants present in this population

and R117H poly-T haplotypes will be considered to determine disease status and will be presented in a table.

- Missing data:
    - As well as looking at the number and proportion of children that have an unclear diagnosis, we will also estimate the number and proportion of children who have missing sweat chloride test data and/or missing genotype and for whom we can therefore not make any statement about the outcome of interest.

Objective 2: Identifying the number of individuals with a designation of CF screen positive, inconclusive diagnosis who have converted to a diagnosis of cystic fibrosis

- Study Design:
    - Study Type: retrospective cohort study
    - Data Source: UK CF Registry data
    - Study Population (eligibility criteria): CFSPID population identified in objective 1
- Outcome:
    - *Outcome:* conversion to diagnosis of CF
    - *Definition of outcome*:
        - Reclassified mutations
        - Increased sweat chloride (>60mmol/L)
        - Clinical diagnosis of CF
    - *Variables used to determine this in the UK CF Registry:*
        - S03sweattestvalue01 – 05
        - S03sweattestdate01 – 05
        - S02genotypemut1
        - S02genotypemut1specify
        - S02genotypemut2
        - S02genotypemut2specify
        - S03DiagnosisMethodPersistantAcuteResporatoryInfection
        - S03DiagnosisMethodOedema
        - S03DiagnosisMethodElectrolyteImbalance

- S03DiagnosisMethodMalnutrition
- S03DiagnosisMethodBronchiectasis
- S03DiagnosisMethodPancreatitis
- S03DiagnosisMethodFertility
- S03DiagnosisMethodLiverDisease
- S03DiagnosisMethodMeconiumIleus
- S03DiagnosisMethodNasalPolyps
- S03DiagnosisMethodRectalProlapse
- Other: birth_year

- <u>Methods:</u>
  - CFTR2 database will be used to determine if any variants have been reclassified and to determine their current status if they have converted from VVCC/VUS to CF-causing which is indicative of a conversion. This will be done using the genotype variables: S02genotypemut1, S02genotypemut1specify, S02genotypemut2, S02genotypemut2specify
  - To consider sweat chloride conversions, longitudinal sweat test data will be analysed using sweat test variables: S03sweattestvalue01 – 05 and S03sweattestdate01 – 05, an increase from below 60mmol/L to over 60mmol/L is considered a conversion
  - Final type of conversion is a clinical conversion which uses the clinical diagnosis methods variables: S03DiagnosisMethodPersistantAcuteRespiratoryInfection, S03DiagnosisMethodOedema, S03DiagnosisMethodElectrolyteImbalance, S03DiagnosisMethodMalnutrition, S03DiagnosisMethodBronchiectasis, S03DiagnosisMethodPancreatitis, S03DiagnosisMethodFertility, S03DiagnosisMethodLiverDisease, S03DiagnosisMethodMeconiumIleus,

S03DiagnosisMethodNasalPolyps, S03DiagnosisMethodRectalProlapse

- Median time to conversion and the mean sweat chloride value for conversion will be calculated
- Clinical conversions will not be included in average time to conversion due to uncertainties when multiple diagnoses were added

- Missing data:
  - A considerable number of individuals may not be able to be considered for a sweat chloride conversion as they may not have more than one sweat test recorded.

Objective 3: Sweat Test Analysis

- Study Design:
  - Study Type: retrospective cohort study
  - Data Source: UK CF Registry data
  - Study Population (eligibility criteria): CFSPID population identified in objective 1
- Outcome:
  - *Outcomes*:
    - Number of sweat tests per individual in CFSPID population
    - Average sweat chloride level at first sweat test
    - Age at age sweat test in CFSPID population
    - Time to next sweat test after the first
    - Proportion of individuals with sweat tests at each age split pre- and post-2014 and split post-2014 by year
  - *Variables used to determine this in the UK CF Registry:*
    - S03sweattestvalue01 – 05
    - S03sweattestdate01 – 05
- Statistical analysis*:*

- All available sweat test data from the variables S03sweattestvalue01 – 05 and S03sweattestdate01 – 05 will be evaluated for the CFSPID population to answer the outcomes above.
- Number of sweat tests per individual will be calculated and the results presented in a bar chart and a table
- The number of individuals with a sweat test in the first 31 days and after the first 31 days will be presented in a table.
- The proportion of sweat tests performed at each age will be presented in both a table and a bar chart.
- The average sweat chloride level across the population at first sweat test will be calculated and the number of individuals with a sweat chloride value 0-10, 10-19, 20-29, 30-39, 40-49, 50-59 (mmol/L) will be presented in a bar chart
- Time to next sweat test will be shown in a histogram
- The proportion of individuals with at least one sweat test at each age will be show in a line diagram and will be split into two graphs. The first will compare pre-CFSPID designation to post-CFSPID in 2014. The second will take the period from 2015-2019 and separate it yearly to determine if there have been any improvements more recently.
- Follow-up after six years will be presented in a table.

- <u>Missing data:</u>
  - The amount of missing sweat test data for individuals will be evaluated considering follow-up added to the registry.
  - Individuals that do not have at least one sweat test will not be able to be included in this analysis.

R packages

The following lists the packages used in R to perform the statistical analysis:

- Ggplot2
- Graphics
- Here

- Hrbrthemes
- Lubridate
- Readr
- Skimr
- Stringr
- Tidyverse
- Virdis

Statistical Analysis Code

Objective 1: Identifying the number of children with CFSPID on the UK CF Registry

Filter complete dataset to produce study population

Study population was filtered to include those diagnosed by newborn screening and those born after 2007

Code

```
Study_Population <- Complete_Data_1 %>% filter
(s03diagnosismethodnewbornscreeni == 1) %>% filter (birth_year > 2007)
```

Filter study population to consider sweat test data

As at least one sweat test is required to classify an individual as CFSPID, the population was filtered to only include those with sweat test data

Code

```
Obj1_Study_Pop <- Study_Population %>% filter (s03sweattestperformed >= 1)
```

Filter for normal and intermediate sweat chloride

First, date variables term was created to group all variables then the study population was mutated to create a new variable that calculates the gap between the first and second sweat test, the first sweat test was used, and the second/third sweat test if it was performed within 31 days of the first. Those with sweat chloride data were filtered into the normal, intermediate, and raised categories.

Code

```
Date.variables <-c ("s03diagnosisdate”, unname
(tidyselect::vars_select(names(Study_Population), starts_with('s03sweattestdate',
ignore.case = TRUE))))

Study_Population <- mutate_at (vars (Date.variables), list (~ as. Date (.,
"%d/%m/%Y"))) %>% mutate (gap.st1.st2=as.numeric (s03sweattestdate02-
s03sweattestdate01))

Normal_SCC <- Obj1_Study_Pop %>% filter (s03sweattestvalue01 < 30 |
(s03sweattestvalue02 < 30 & gap.st1.st2 < 31))

Intermediate_SCC <- Obj1_Study_Pop %>% filter (between
(s03sweattestvalue01,30,59) | (between(s03sweattestvalue02,30,59) & gap.st1.st2
< 31))

Raised_SCC <- Obj1_Study_Pop %>% filter (s03sweattestvalue01 >= 60 |
(s03sweattestvalue02 >= 60 & gap.st1.st2 < 31))
```

CFSPID Definition 1: normal SCC & 2 CFTR variants

The normal sweat chloride population was then filtered for those with a genotype that fits a diagnosis of CFSPID. Genotype information was retrieved from CFTR2 database to classify as VVCC or VUS.

Code

```
CFSPID_Def_1 <- Normal_SCC %>% filter (str_detect (s02genotypemut1,
'\\b(R117H) \\b') | str_detect (s02genotypemut2, '\\b(R117H|D1152H|T1246I)
\\b') | str_detect (s02genotypemut1specify,'\\b (Cys491Phe|M952I) \\b') |
str_detect (s02genotypemut2specify, '\\b
(TG12TS|Q237E|1523delCTT|3717|E1124del|R117H) \\b') | str_detect
(regid_anon, '\\b(3IUSCAS) \\b'))
```

CFSPID Definition 2: intermediate SCC & 0/1 CFTR variants

The intermediate sweat chloride population was filtered for those with a genotype that fits a diagnosis of CFSPID, and genotype information was also retrieved from CFTR2 database to classify as VVCC or VUS.

Code

```
CFSPID_Def_2 <- Intermediate_SCC %>% filter (str_detect (s02genotypemut1, '\\b(1161delC|A455E|P67L|R117H) \\b') | str_detect (s02genotypemut2, '\\b(2insA|D1152H|D443Y|F575Y|R1070W|T1246I|R117H) \\b') | str_detect (s02genotypemut2specify, '\\b (L375F|c.1384G|p.Ser118Phe|490|c.4339del|R117H9T|Gln1291His|709C>G|473G >A|GIN237Glu|Y1073C) \\b'))
```

Consider reversed diagnosis

The study population was filtered for those with a diagnosis reversal and then reasons for the reversal were considered. Those with CFSPID were filtered out and added to the CFSPID population.

Code

```
Study_Population_DR_Y <- Study_Population_DR %>% filter (s03diagnosisreversed == "Y")

CFSPID_DR <- Study_Population_DR_Y %>% filter (str_detect (s03diagnosisreversedreasonotherd, '\\b (CFSPID|SPID) \\b'))
```

R117H Analysis

R117H analysis was undertaken to determine those with a CF-causing haplotype to exclude them from the CFSPID population. The amount of R117H data present on the registry was also noted.

Code

```
R117H <- CFSPID_Population %>% filter (str_detect (s02genotypemut1, '\\b(R117H) \\b') |str_detect (s02genotypemut1specify, '\\b(R117H) \\b') |str_detect (s02genotypemut2, '\\b(R117H) \\b') |str_detect (s02genotypemut2specify, '\\b(R117H) \\b'))

R117H_5T <- R117H %>% filter (s02genotypemut2polyttract5t == 1|s02genotypemut2polyttract5t == 1)
```

Combining to form CFSPID population

The potential CFSPID individuals from definition one, definition two and diagnosis reversal were combined. Then, the CFSPID population was screened to remove individuals with R117H;5T to result in a final population.

Code

```
CFSPID_Population_1 <- rbind (CFSPID_Def_1, CFSPID_Def_2, CFSPID_DR)

CFSPID_Population <- anti_join (CFSPID_Population_1, R117H_5T)
```

Objective 2: Identifying the number of individuals with a designation of CF screen positive, inconclusive diagnosis that have converted to a diagnosis of cystic fibrosis

Filter for individuals with more than one sweat test

To help determine sweat chloride conversions, the CFSPID population were filtered to include those with more than one sweat test. Another sweat test at a later date is required to indicate a conversion has occurred.

Code

```
Obj2_All_SCC <- CFSPID_Population %>% filter (s03sweattestperformed > 1)
```

Filter to detect any individuals with subsequent sweat test >60mmol/L

The population were further filtered to consider those with a second or third sweat chloride concentration that is greater than the diagnostic threshold. The two populations were combined, and any duplicates were removed to produce a final group to analyse.

Code

```
Obj2_SCC_1 <- Obj2_All_SCC %>% filter (s03sweattestvalue02 >= 60)

Obj2_SCC_2 <- Obj2_All_SCC %>% filter (s03sweattestvalue03 >= 60)

Obj2_SCC_Dup <- rbind (Obj2_SCC_1, Obj2_SCC_2)

Obj2_SCC <- Obj2_SCC_Dup [! duplicated (Obj2_SCC_Dup$regid_anon),]
```

Detect any reclassified mutations

The study population was screened to determine if any reclassified mutations are present which were identified from the CFTR2 database. The CFSPID population were screened for these reclassified mutations and if they were present

Code

```
Obj2_Reclassifed <- CFSPID_Population %>% filter (str_detect (s02genotypemut1, '\\b(Y569D|H199Y) \\b') | str_detect (s02genotypemut1specify, '\\b(Y569D|H199Y) \\b') | str_detect (s02genotypemut2, '\\b(Y569D|H199Y) \\b') |str_detect (s02genotypemut2specify, '\\b(Y569D|H199Y) \\b'))
```

Detect any clinical conversions

The study population had their diagnosis methods considered in more detail to see if any individual has multiple methods of diagnosis. These individuals had their clinical picture evaluated to determine whether the additional methods of diagnosis may have been added subsequently.

Code

```
Obj2_Clinical <- CFSPID_Population %>% filter
(S03DiagnosisMethodPersistantAcuteRespiratoryInfection == 1 |
S03DiagnosisMethodOedema == 1 |S03DiagnosisMethodElectrolyteImbalance == 1
| S03DiagnosisMethodMalnutrition == 1 | S03DiagnosisMethodBronchiectasis == 1
| S03DiagnosisMethodPancreatitis == 1 | S03DiagnosisMethodFertility == 1 |
S03DiagnosisMethodLiverDisease == 1 | S03DiagnosisMethodMeconiumIleus == 1 |
S03DiagnosisMethodNasalPolyps == 1 | S03DiagnosisMethodRectalProlapse == 1)
```

Combine reclassified mutations and sweat chloride conversions

Reclassified mutations, sweat chloride conversions and clinical conversions were combined to form a group of conversions.

Code

```
Obj2_Study_Pop <- rbind (Obj2_SCC, Obj2_Reclassifed, Obj2_Clinical)
```

Objective 3: Sweat Test Analysis

Number of sweat tests in CFSPID population

The CFSPID population were filtered to determine the number of sweat tests performed on individuals in this group. This was then presented graphically in a bar chart.

Code

```
SweatPer1 <- CFSPID_Population %>% filter (s03sweattestperformed == 1)

SweatPer2 <- CFSPID_Population %>% filter (s03sweattestperformed == 2)

SweatPer3 <- CFSPID_Population %>% filter (s03sweattestperformed == 3)

SweatPer4 <- CFSPID_Population %>% filter (s03sweattestperformed == 4)
SweatPer5 <- CFSPID_Population %>% filter (s03sweattestperformed == 5)
```

Graph

```
Number_of_sweat_tests <- c (" 1 sweat test only", "2 sweat tests", "3 sweat tests",
"4 sweat tests", "5 sweat tests")

Number_of_individuals <- c (96, 70, 20, 6, 1)

Combine <- data. frame (Number_of_sweat_tests, Number_of_individuals)

ggplot (Combine, aes (x= Number_of_sweat_tests, y= Number_of_individuals)) +
geom_bar (stat = "identity")
```

Multiple sweat tests within first 31 days

The variable 'gap.st1.st2' was created to determine the gap between the first and second recorded sweat tests. This variable was then used to filter those another sweat test within 31 days of the first.

Code

```
SCC1_2_31d_CFSPID <- CFSPID_Population %>% filter (gap.st1.st2 <31)
```

Mean sweat test value

The mean and distribution of sweat chloride concentrations in the CFSPID population were analysed to determine whether it is skewed towards the normal or the intermediate classification.

Graph

```
SC_level <- c ("0-10", "10-19","20-29", "30-39", "40-49", "50-59")

Number <- c (0,26,70,50,27,16)

Stratify <- data. frame (SC_level, Number)
```

```
ggplot (Stratify, aes (x=SC_level, y=Number)) + geom_bar (stat = "identity") +
theme_bw ()
```

Age at each sweat test

The variable age.st.1 was created for each sweat test one to five which determines the age at which each sweat test was performed. The CFSPID population was filtered to calculate the proportion of individuals at each age with at least one sweat test performed.

Code

```
Age1_ST <- CFSPID_Population %>% filter (age.st.1 < 1)

Age2_ST <- CFSPID_Population %>% filter (age.st.1 < 2)

Age3_ST <- CFSPID_Population %>% filter (age.st.1 < 3)

Age4_ST <- CFSPID_Population %>% filter (age.st.1 < 4)

Age5_ST <- CFSPID_Population %>% filter (age.st.1 < 5)

Over5 <- CFSPID_Population %>% filter (age.st.1 > 5)
```

Graph

```
Age_sweat <- c("<1", "<2", "<3", "<4", "<5", ">5","<1", "<2", "<3", "<4", "<5", ">5",
"<1", "<2", "<3", "<4", "<5", ">5", "<1", "<2", "<3", "<4", "<5", ">5", "<1", "<2", "<3",
"<4", "<5", ">5")

Proportion_sweat <- c(rep("ST1",6), rep("ST2",6), rep("ST3",6), rep("ST4",6),
rep("ST5",6))
```

Proportion_value <- c
(0.91,0.06,0.02,0.0,0.01,0.0,0.49,0.24,0.05,0.06,0.04,0.12,0.22,0.15,0.07,0.15,0.15,
0.26,0.29,0.28,0.14,0.15,0,0.14,0,0,0,0,0,1)

Time_graph <- data. frame (Age_sweat, Proportion_sweat, Proportion_value)

ggplot (time_graph, aes (fill= Age_sweat, x=Proportion_sweat,
y=Proportion_value)) + geom_bar (position = "stack", stat = "identity")

Average time to next sweat test after first 30 days

The gap.st1.st2 and other similar variables: gap.st2.st3, gap.st3.st4 were used to calculate the gap between the first and second sweat test along with the second and third and third and fourth. The first 30 days were excluded as these may have been in the same clinical visit and that would not give a true reflection of follow-up.

Code

ST1_2_CFSPID <- filter (gap.st1.st2 >= 30)

ST2_3_CFSPID <- filter (gap.st2.st3 >= 30)

ST3_4_CFSPID <- filter (gap.st3.st4 >= 30)

Graph

The variable 'time.to next.st.after30days' was created to calculate the gap between sweat tests and was used to make a histogram showing time to follow up.

hist (CFSPID_Population$time.to.next.st.after30days)

Consistency of follow-up

The proportion of individuals with at least one sweat test at each age first in the period 2008-2014 and then 2015-2019 was considered to determine whether follow up has improved after the introduction of the CFSPID designation.

The CFSPID population was split into two populations and number of sweat tests performed was calculated. The numbers in the CFSPID population had to be calculated carefully in the 2015-2019 group as the younger individuals did not have a full length of follow-up. A proportion was then calculated and shown graphically in a line chart.

Graph

```
CFSPID.2008.2014 <- CFSPID_Population %>% filter (between (birth_year, 2008, 2014))

CFSPID.2015.2020 <- CFSPID_Population %>% filter (between (birth_year, 2015, 2019))

Birth <- c (rep ("2008-2014",6), rep ("2015-2019",5))

Age <- c ("<1", "1-2", "2-3", "3-4", "4-5", "5<","<1", "1-2", "2-3", "3-4", "4-5")

Number_CFSPID <- c (107,107,107,107,107,107,83,83,69,49,36)

At_least_1_ST <- c (95,21,7,6,7,13,78,17,4,5,2)

proportion <- c (0.88,0.18,0.068,0.058,0.068,0.15,0.91,0.25,0.020,0.083,0)

Data <- data.frame(birth, age, proportion)

ggplot (Data, aes (x=age, y=proportion, group=birth, colour=birth)) + geom_point () + geom_line () + theme_bw ()
```

Graph

The 2015-2019 population was then split into individual years to see if more recently follow-up has been improving.

```
Birth <- c (rep ("2008-2014",6), rep ("2014", 6), rep ("2015",5), rep ("2016",4), rep ("2017",3), rep ("2018",2), rep ("2019",1))

Proportion <- c (0.88,0.18,0.068,0.058,0.068,0.15,0.94,0.22,0.1,0.1,0.056,0.17,0.89,0.21,0,0.053,0,0.74,0.21,0,0.21,1,0.077,0.23,1,0.45,1)

Age <- c("1", "1-2", "2-3","3-4","4-5","5>","1", "1-2", "2-3","3-4","4-5","5>","1","1-2","2-3","3-4", "4-5","1","1-2","2-3","3-4", "1","1-2","2-3", "1", "1-2", "1")

Follow_up <- data. frame (birth, proportion, age)

ggplot (Follow_up, aes (x=Age, y=Proportion, group=Birth, colour=Birth)) +geom_point () + geom_line () + theme_bw ()
```

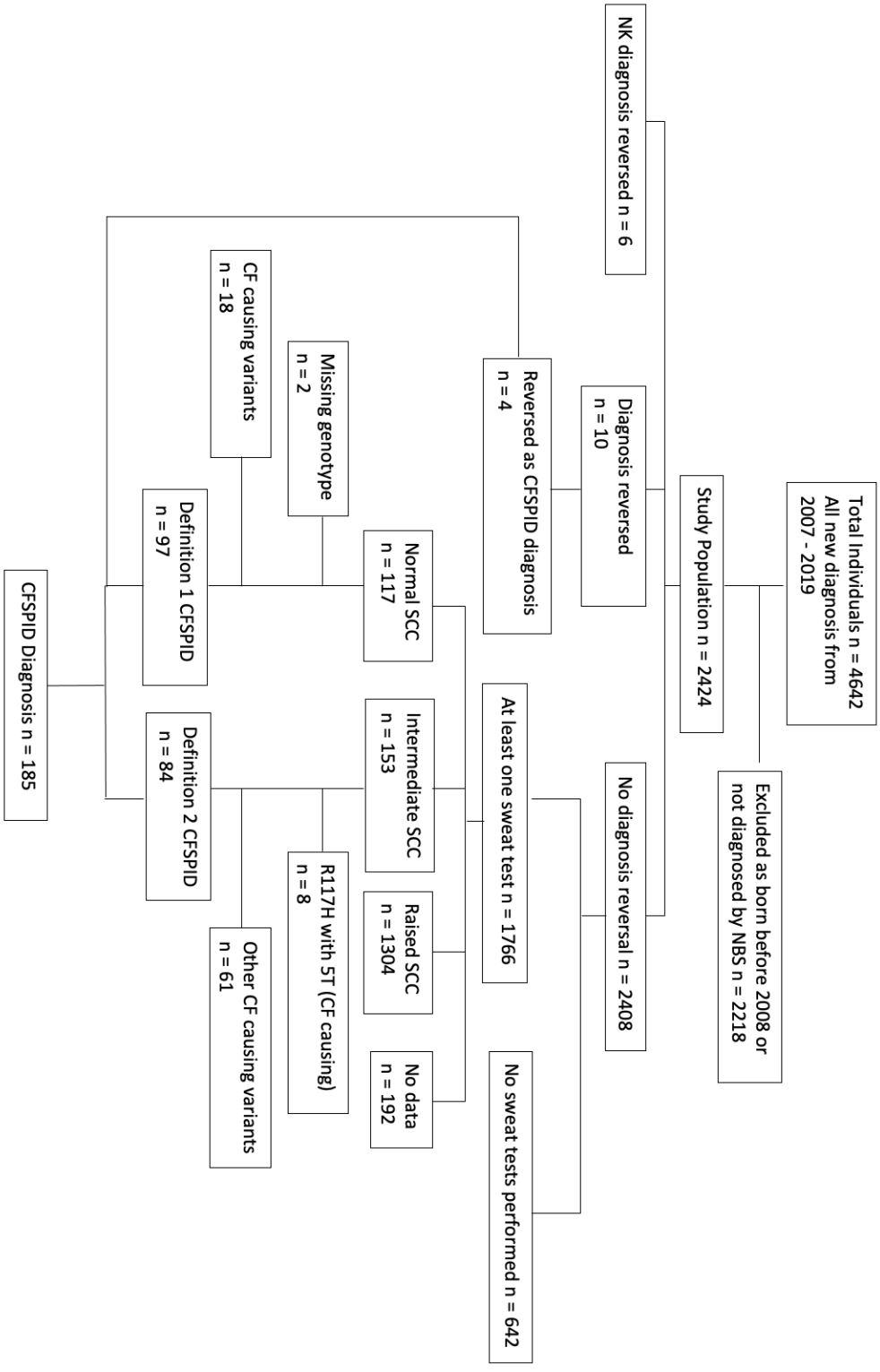
Follow up at age 6 and into adolescence

Follow up greater than six was calculated using the age at sweat test variable and the CFSPID population was filtered to include those who had a sweat test when they were over six.

Code

```
Over6 <- CFSPID_Population %>% filter (age.st.1 >= 6 | age.st.2 >= 6| age.st.3 >= 6| age.st.4 >= 6| age.st.5 >= 6)
```

CONSORT Diagram



CFTR1 Cases of variants of unknown significance (VUS)

There were certain variants present in the CFTR1 database that were not present in the CFTR2 database. The CFTR1 database contain individual case studies on variants and clinicians often provide the clinical picture of an individual with this variant. The case studies and clinical pictures present on the CFTR1 database were compared to the individuals in our CFSPID population to see if any similarities could be detected.

M952I (c.2856G>C)

This variant was present in 2 patients on the CFTR1 database (66). One patient presented with congenital bilateral absence of vas deferens (CBAVD), a common CFTR related disorder (CFTR-RD) with their other variant being R117H, a common variant of varying clinical consequence.

The other individual with this variant has severe lung disease, pancreatic insufficiency, a sweat test of 95mmol/L and no other identified variant (66).

P1013L (c.3038C>T)

This variant was also found in two individuals on the CFTR1 database (67). The first patient has this P1013L in combination with 2752-15C>G, a variant which is not present on either database. This individual one sweat test reported as 65mmol/L which is considered raised.

A second patient with P1013L presented with moderate lung disease, pancreatic sufficiency, negative sweat tests and infertility (67). This is more of a mixed clinical picture compared to the first patient.

In our population, the sweat chloride value was in the normal sweat chloride range 11mmol/L and the presence of clinical symptoms is unknown, and it is not possible to conclude the classification of either variant and there are no direct similarities between our individual and the case studies.

L375F (c.1125A>C)

The CFTR1 database also contained 2 case studies for the L375F variant (68). One patient, with L375F in combination with R117H, presented with CBAVD, pancreatic sufficiency, moderate lung disease and a negative sweat test.

The second patient, whose other variant was G551D (CF-causing) had a similar picture but presented with congenital unilateral absence of vas deferens (CUAVD), hydrocele, pancreatic sufficiency and no lung disease (68).

It was thought that having a CF-causing variant, as the second patient does, would result in a more severe clinical picture however in these case studies, the VVCC in combination with L375F has resulted in a more severe phenotype.

This variant was seen in the CFSPID population of this study in combination with a CF-causing variant. Both a normal and intermediate sweat chloride was noted and there are some similarities between these case studies as none have a conclusive diagnosis of CF or a positive sweat test.

S158N (c.473G>A)

This variant is present once in the CFTR1 database (69). It was a heterozygous variant seen in a patient with chronic pancreatitis. A second variant was investigated, and no other common variants were detected apart from two sequence variations of the CFTR gene.

This variant was present in the CFSPID population of this study in combination with a CF-causing variant and sweat tests were in the intermediate range. This differs from the case study above as a second CF-causing variant (F508del) was identified.

Y1073C (c.3218A>G)

This variant was present on the CFTR1 database once in a patient diagnosed with neonatal hypertrypsinemia at 6 months (70). Their sweat chloride concentration

was between 36-44 mmol/L, and they were pancreatic sufficient with no lung disease. No second CFTR variants was detected in this individual. This is similar to the CFSPID population as the majority are pancreatic sufficient and do not have any lung disease (18).

This variant was present once in our CFSPID population in combination with G542X, which is CF-causing. This individual had sweat tests in a similar range to the patient in the case study in the intermediate sweat chloride category.

Overall, the significance of these variants is still unknown, and decisions about the classification of variants cannot be made from a handful of case studies. Further case studies and information will be required in order to correctly classify these very rare variants.

Clinical conversions to a diagnosis of CF

Persistent acute respiratory infection

All individuals that converted due to persistent acute respiratory infection had multiple sweat tests performed and one individual fit both sweat chloride and clinical symptom criteria for conversion.

Persistent acute respiratory infection was also seen alongside meconium ileus, and it is unclear how it fits with the clinical picture as meconium ileus usually occurs in the neonatal period and is indicative of CF. This circumstance may have been considered CF from the neonatal period even though their genotype and sweat chloride concentration suggest CFSPID.

Furthermore, persistent acute respiratory infection was also seen in combination with electrolyte imbalance. Electrolyte disturbance often occurs at birth in infants with CF, for example, hyponatraemic dehydration and this could lead to a quick diagnosis of CF (2). The combination of clinical symptoms does suggest a diagnosis of CF even on a background of CFSPID.

Malnutrition

Malnutrition was seen as a diagnosis method in the CFSPID population and can occur if cystic fibrosis is left untreated in the neonatal period (34). This also suggests a diagnosis of CF regardless of the CFSPID genotype and sweat chloride concentration.

Liver disease

Another method of diagnosis present alongside newborn screening was liver disease, which presents later in life than some of the methods above.