



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/178840/>

Version: Published Version

Article:

Totton, N., Bradburn, M., Hoo, Z.H. et al. (2021) Prospectively predicting *Pseudomonas aeruginosa* infection/s using routine data from the UK cystic fibrosis register. *Health Science Reports*, 4 (4). e381. ISSN: 2398-8835

<https://doi.org/10.1002/hsr2.381>

Reuse



This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Prospectively predicting *Pseudomonas aeruginosa* infection/s using routine data from the UK cystic fibrosis register

Nikki Totton¹  | Mike Bradburn¹ | Zhe Hui Hoo^{1,2}  | Jen Lewis¹ | Daniel Hind¹ | Carla Girling¹ | Elizabeth Shepherd³ | Julia Nightingale³ | Thomas Daniels³ | Jane Dewar⁴ | Sophie Dawson⁴ | Mary Carroll³ | Mark Allenby³ | Frank Edenborough² | Rachael Curley² | Charlotte Carolan² | Martin Wildman²

¹Clinical Trials Research Unit, School of Health and Related Research, University of Sheffield, Sheffield, UK

²Sheffield Adult Cystic Fibrosis Centre, Sheffield Teaching Hospital NHS Foundation Trust, Sheffield, UK

³Department of Adult Cystic Fibrosis, University Hospital Southampton NHS Foundation Trust, Southampton, UK

⁴Wolfson Cystic Fibrosis Centre, Nottingham University Hospital NHS Trust, Nottingham, UK

Correspondence

Martin Wildman, Sheffield Adult CF Centre, Northern General Hospital, Herries Road, Sheffield, S5 7AU, UK.
Email: martin.wildman3@nhs.net

Funding information

NHS England Commissioning for Quality and Innovation

Abstract

Rationale and aims: Lung health of people with cystic fibrosis (PwCF) can be preserved by daily use of inhaled therapy. Adherence to inhaled therapy, therefore, provides an important process measure to understand the success of care and can be used as a quality indicator. Defining adherence is problematic, however, since the number of prescribed treatments varies considerably between PwCF. The problem is less pronounced among those with *Pseudomonas aeruginosa* (PA), for whom at least three daily doses of nebulized therapy should be prescribed and who thus constitute a more homogeneous group. The UK CF Registry provides routine data on PA status, but data are only available 12 months after collection. In this study, we aim to prospectively identify contemporary PA status from historic registry data.

Method: UK CF Registry data from 2011 to 2015 for PwCF aged ≥ 16 was used to determine a pragmatic prediction rule for identifying contemporary PA status using historic registry data. Accuracy of three different prediction rules was assessed using the positive predictive value (PPV). The number and proportion of adults predicted to have PA infection were determined overall and per center for the selected prediction rule. Known characteristics linked to PA status were explored to ensure the robustness of the prediction rule.

Results: Having CF Registry defined chronic PA status in the two previous years is the selected definition to predict a patient will have PA infection within the current year (population-level PPV = 96%-97%, centre level PPV = 85%-100%). This approach provides a subset of data between 1852 and 1872 patients overall and a range of 8 to 279 patients per center.

Conclusion: Historic registry data can be used to contemporaneously identify a subgroup of patients with chronic PA. Since this patient group has a narrower treatment schedule, this can facilitate a better benchmarking of adherence across centers.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Health Science Reports* published by Wiley Periodicals LLC.

KEYWORDS

adherence, cystic fibrosis, outcome, process, quality improvement

1 | INTRODUCTION

Cystic fibrosis (CF) is an archetypal long-term condition for which there is as yet no cure, but regular use of preventative therapies can improve health outcomes by reducing the frequency of exacerbations and attenuating lung function (forced expiratory volume in 1 second, FEV₁) decline.¹ However, adherence to preventative inhaled treatments in CF is generally low (30%-35%) and low adherence is associated with worse health outcomes.^{2,3} Measuring adherence at an individual level is important for treatment planning and diagnosis in the event of declining lung health. Measuring system-level adherence is also important because benchmarking across centers within a community of practice can provide a basis to share strategies for improvement. Various quality improvement (QI) initiatives in CF have transformed the delivery of healthcare, for example, streamlined approaches to managing acute exacerbations and increased prescription of efficacious preventative inhaled therapies.⁴ Improving adherence to inhaled therapies is, therefore, a logically important target for QI projects. QI projects focused on adherence can compare between and within centers, therefore providing an understanding of the variation of care across UK centers.

System-level indicators are most useful when they are independent of local constraints on data capture since the factors that impact quality of care may also impact data capture. Data-logging nebulizers available in CF can objectively and accurately record each dose of inhaled treatments taken by patients every day.⁵ However, accurate prescriptions data may be unavailable, which impacts the ability to calculate percentage adherence. Yet adherence among people with CF is typically presented as a percentage because the required number of treatment doses will vary according to disease severity. A person without chronic PA and FEV₁ 80% may only require dornase alfa once daily, whereas someone with chronic PA, FEV₁ 30% with large volume of sputum may require alternating antibiotics twice daily, dornase alfa once daily, and hypertonic saline twice daily.

Identifying a subgroup of people with CF (PwCF) where the prescription, and therefore denominator for adherence, can be implied using routine data, despite local challenges in data capture, is an important step toward providing adherence comparisons at the center level. One option is to focus on those who have chronic *P. aeruginosa* (PA) infection as consensus exists on the treatment these patients should receive. All major CF guidelines recommend the use of long-term inhaled antibiotics and mucolytics for people with chronic PA infection, and this provides a minimum denominator of three drugs for those PwCF with chronic PA. Using registry data to identify patients with chronic PA and designating that these patients should receive at least three daily doses of inhaled drugs (denominator) allows adherence data captured by CFHealthHub⁶ (which provides numerator data) to calculate an adherence rate even when

prescription data are missing. Although these data will be less accurate in centers without accurate prescription data, this method of calculating adherence data does allow a standardized approach to be used in a well-defined subset of PwCF allowing center benchmarking that can start an informed exploration of practice.⁷

The aim of system-level indicators using routine data where possible allows all centers to be included within comparisons, even those that do not have the resources to provide data consistently as the identification of the subgroup with PA uses routine data from the CF Registry. Published UK CF Registry provides historic data on PA status 1 year in arrears. As PA status may fluctuate over time, it is important that the contemporary status is understood as accurately as possible to identify the PA subgroup. Thus, the overarching aim of this paper is to develop a rule to accurately predict which patients will have PA in the current year on the basis of previous years' data collected routinely in the UK CF Registry. It is important to ensure that any prediction rule has an appropriate compromise between sensitivity and specificity to identify a sufficient number of people to allow center comparisons. Thus, the analysis described in this paper sought to determine the optimal rule for prospectively identifying patients who have PA in the current year (trading off accuracy and number). In addition, we explored how other patient characteristics that might be expected to be associated with PA status such as age were related to the PA identifying strategies used.

2 | METHODS

Retrospective data were collected from the UK CF Registry database for annual reviews between 2011 and 2015 from all UK CF centers. National Health Service (NHS) research ethics approval was granted for the use of UK CF Registry data (Huntingdon Research Ethics Committee 07/Q0104/2), and under these terms, the UK CF Trust Steering Committee approved this study. The data received had pseudo-anonymized patient and center details.

The following data were obtained from the registry:

- **Demographics:** CF center identifier, gender, age
- **Body mass index (BMI)** in kg/m²
- **FEV₁** during annual review (in % predicted, calculated with GLI equations⁸)
- **Indices of multiple deprivation (IMD)** collected in each of the four UK countries and adjusted to the equivalent English value⁹
- ***P. aeruginosa* status:** chronic (registry definition of three or more positive samples in a 1-year period, regardless of the number of samples sent), intermittent (one or two positive samples in a 1-year period), none (no positive samples identified in a 1-year period)

- **Pancreatic status:** pancreatic insufficient (those on pancreatic replacement therapy [PERT]) or pancreatic sufficient (those not on PERT)
- **CF-related diabetes (CFRD):** present or not present, as defined by the UK CF Trust guideline¹⁰
- **Annual total intravenous (IV) antibiotic use:** total annual number of home and hospital IV antibiotic days
- **Lung transplant:** year of transplant

All adults aged 16 years and above were included in the analyses, with the exception of those undergoing lung transplantation who were excluded from the year of transplant onwards.¹¹

For this analysis, we compared the accuracy of three prediction rules that have been selected from clinical knowledge of factors affecting a patient's PA status, and that can be derived from routinely collected data in the CF Registry:

1. Patients with three or more positive PA samples in the previous year (these are classified as “chronic” according to the UK CF Registry guidance),
2. Patients with any positive PA samples in the previous year (“chronic” as above plus patients with one or two PA samples that are classified as “intermittent” according to the UK CF Registry guidance),
3. Patients with any positive PA samples and/or at least 14 days of IV antibiotics (home or hospital) in the previous year.

The rules aim to predict whether a patient will have any positive PA samples within the current year (ie, been classified as “chronic” or “intermittent” by the CF Registry).

The accuracy of the prediction rule was assessed using the positive predictive value (PPV) of the prediction against the outcome of the patients having positive PA samples within the current year based on past years' data. The definition producing the highest PPV (or equivalently, the smallest proportion of patients misclassified as

having PA) was taken on to further analysis in which we assessed the impact of increasing the number of years of past data to include in the prediction rule. We calculated the PPV for one, two, three, and four consecutive years' PA diagnosis in order to find the best compromise between accuracy and number of patients included in the subgroup.

Once a final prediction rule had been selected, we used logistic regression to assess the robustness of the prediction rule by calculating the PPV adjusted for other patient characteristics. The covariates included in the model were as follows: age, pancreatic status, FEV₁, BMI, presence of CFRD, IMD, and IV days all as fixed effects, along with center as a random effect. The functional form of continuous covariates was assessed visually using lowess smoothing¹² and using fractional polynomials¹³; this is particularly important for age, since the relationship has been shown to be nonlinear within younger patients.¹⁴

As multiple years' data are available to make predictions, the model was repeated with all data possible, allowing the consistency of the covariates' effect on the prediction to be assessed. The discriminant ability of the model was quantified using the area under the receiver-operator curve. Since most patients contributed to more than 1 year, the findings of separate predictive models were not combined.

Funnel plots¹⁵ were created to show the PPV by center and identify any outliers from the prediction rule, using both unadjusted and adjusted PPV. No imputation was performed on missing data. All analyses were performed using R V3.4.1¹⁶ and Stata V15.¹⁷

3 | RESULTS

The flowchart in Figure 1 shows the number of patients contributing data to the UK CF Registry in each year. One of the 28 centers did not contribute data to years 2012 and 2013, this center was excluded from all center-level summaries. The characteristics for the population of patients included in the final master dataset are shown in Table 1, split by year.

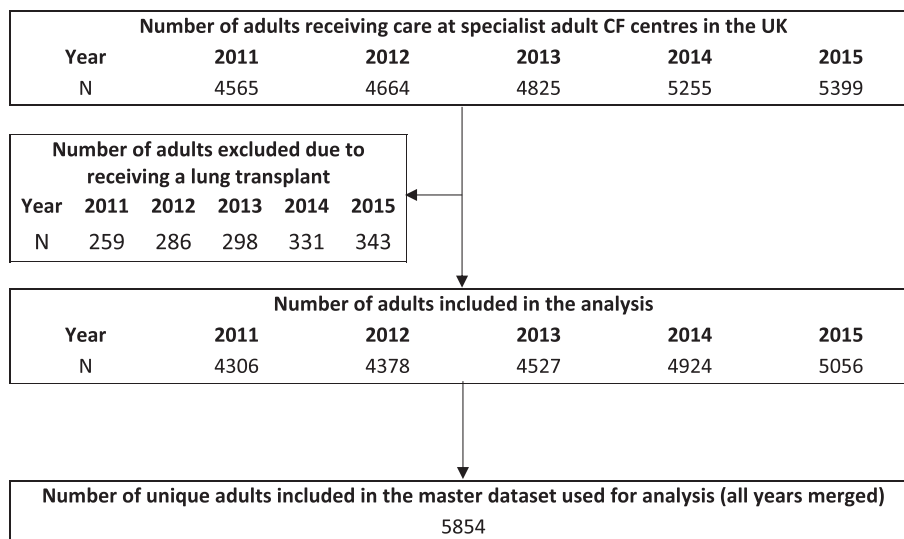


FIGURE 1 Flowchart of patients within the final dataset

TABLE 1 Year-by-year patient characteristics

	2011 N = 4306	2012 N = 4378	2013 N = 4527	2014 N = 4924	2015 N = 5056
Age (years)					
Mean (SD)	29.3 (10.1)	29.7 (10.3)	30.0 (10.4)	30.4 (10.6)	30.6 (10.7)
BMI (kg/m ²)					
Mean (SD)	22.6 (3.8)	22.5 (3.8)	22.6 (3.9)	22.7 (3.9)	22.9 (4.1)
Missing (%)	74 (2%)	136 (3%)	125 (3%)	111 (2%)	124 (2%)
CFRD N (%)					
Yes	1271 (30%)	1342 (31%)	1421 (31%)	1601 (33%)	1676 (33%)
No	3035 (70%)	3036 (69%)	3106 (69%)	3323 (67%)	3380 (67%)
FEV ₁ (% predicted)					
Mean (SD)	62.9 (23.6)	62.0 (23.9)	62.5 (23.8)	63.4 (23.9)	63.4 (23.8)
Missing (%)	110 (2%)	214 (5%)	180 (4%)	200 (4%)	208 (4%)
IV days					
Mean (SD)	25.0 (36.6)	24.7 (38.1)	23.9 (35.8)	23.4 (37.0)	22.0 (35.3)
Missing (%)	58 (1%)	5 (<1%)	5 (<1%)	0 (0%)	0 (0%)
IMD					
Mean (SD)	22.0 (15.7)	21.6 (15.4)	21.7 (15.4)	21.9 (15.6)	22.0 (15.6)
Quintile 1 – least deprived	764 (18%)	795 (18%)	827 (18%)	875 (18%)	898 (18%)
Quintile 2	828 (19%)	853 (19%)	880 (19%)	946 (19%)	981 (19%)
Quintile 3	851 (20%)	875 (20%)	894 (20%)	975 (20%)	999 (20%)
Quintile 4	810 (19%)	822 (19%)	850 (19%)	948 (19%)	956 (19%)
Quintile 5 – most deprived	749 (17%)	735 (17%)	765 (17%)	842 (17%)	879 (17%)
Missing (%)	304 (7%)	298 (7%)	311 (7%)	338 (7%)	343 (7%)
Sex N (%)					
Female	1926 (45%)	1969 (45%)	2053 (45%)	2223 (45%)	2297 (46%)
Male	2380 (55%)	2409 (55%)	2474 (55%)	2701 (55%)	2759 (54%)
<i>P. aeruginosa</i> status N (%)					
Chronic	2447 (57%)	2457 (56%)	2428 (54%)	2477 (50%)	2432 (48%)
Intermittent	532 (13%)	527 (12%)	605 (13%)	726 (15%)	709 (14%)
None	1327 (31%)	1394 (32%)	1494 (33%)	1721 (35%)	1915 (38%)
Pancreatic status N (%)					
Insufficient	3508 (81%)	3574 (82%)	3682 (81%)	4008 (81%)	4107 (81%)
Sufficient	699 (16%)	727 (17%)	774 (17%)	869 (18%)	886 (18%)
Missing	99 (2%)	77 (2%)	71 (2%)	47 (1%)	63 (1%)

Abbreviations: BMI, body mass index; CFRD, cystic fibrosis related diabetes; FEV₁, forced expiratory volume; IMD, indices of multiple deprivation; IV, intravenous.

For each of the four time periods, 94% of patients with chronic PA were found to have PA in the subsequent year. The other prediction rules included a greater number of patients but a smaller proportion with PA (range 79%-90%; Table 2).

Using the prediction rule of chronic PA status in previous years, the optimum number of consecutive years' data to use was explored and the results are shown in Table 3. As expected, the PPV increases with extra years' data (range: 94%-98%), and Figure 2 shows this against the number within the subset, which decreases as more years'

data are included. Taking the two together suggests the optimal threshold for the prediction rule is to select those with two successive years of a chronic PA status to predict the current years' status, thereby allowing a sufficiently large sample to be used for future research (average N = 1862) without appreciably reducing the confidence in the accuracy of the diagnosis (average PPV = 96%). The contingency table for the PPV calculations of this chosen prediction rule for 2013 is shown in Table 4, and the contingency tables for 2014 and 2015 are in Table A1 and the resulting funnel plots in Figure A1.

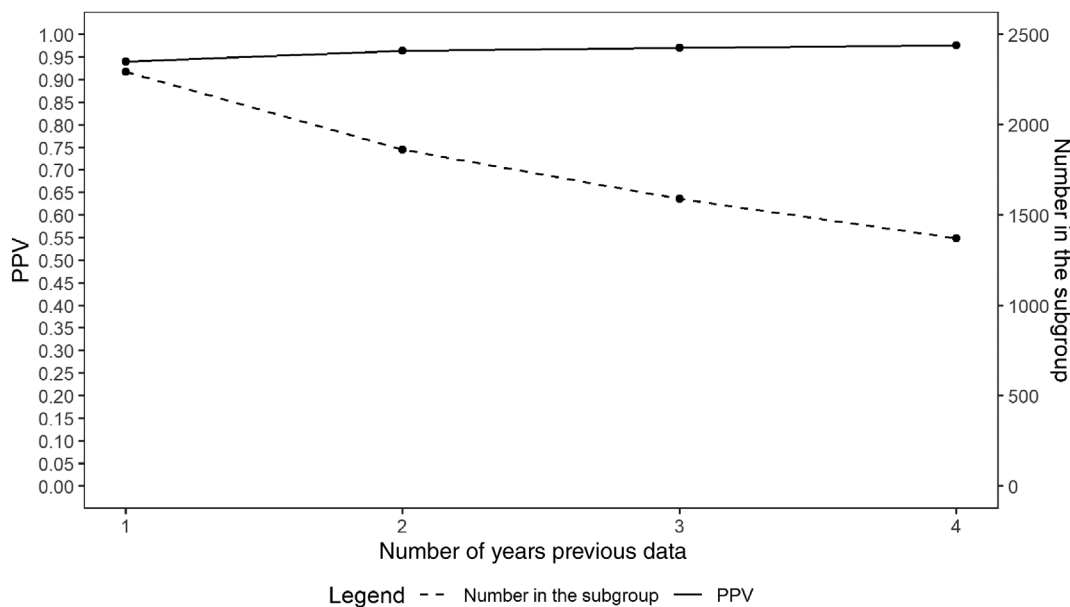
TABLE 2 PPV for three initial prediction rules

Prediction rule	Predicted year; PPV (%)			
	2012 N = 3931	2013 N = 4093	2014 N = 4284	2015 N = 4630
Chronic PA status	94% (93%, 95%)	94% (93%, 95%)	94% (93%, 95%)	94% (94%, 95%)
Chronic or intermittent PA status	90% (89%, 91%)	90% (89%, 91%)	88% (87%, 89%)	87% (86%, 88%)
Chronic or intermittent PA status or ≥ 14 IV days	84% (83%, 86%)	83% (82%, 85%)	81% (79%, 82%)	79% (78%, 81%)

Abbreviations: IV, intravenous; PA, *Pseudomonas aeruginosa*; PPV, positive predictive value.

TABLE 3 Positive predictive value (PPV) for four different years' data in the prediction rule

Number of consecutive previous years	Predicted year; PPV (%)			
	2012	2013	2014	2015
1 year	94% (93%, 95%)	94% (93%, 95%)	94% (93%, 95%)	94% (94%, 95%)
2 years	-	97% (96%, 97%)	96% (96%, 97%)	96% (95%, 97%)
3 years	-	-	97% (96%, 98%)	97% (96%, 98%)
4 years	-	-	-	98% (97%, 98%)

**FIGURE 2** Trade-off between PPV and number within the subset for each number of years' data (averaged across all years)

3.1 | Center comparisons

The PPV was reasonably consistent across the 27 centers (range: 85%–100%). The number selected in the subgroup for each center had a median of at least 49 (from 2013 prediction year) with a range from 8 to 279 (over all prediction years). More details are in Table A2.

3.2 | Predictive modeling

The final analyses sought to identify whether the PPV were associated with the patient's demographic and clinical features. These

analyses included patients that were recorded as having PA in two successive years and were aged under 60. Patients over 60 are both uncommon and also much more likely to be pancreatic sufficient than other age groups, suggesting that they have a preponderance of less severe phenotypes. Analyses were undertaken separately for patients in 2013, 2014, and 2015 separately (Table 5).

The models for three different years showed no consistency for significant predictors, and only age appeared more than once. In all years, the predictive ability of the model was low with an area under the curve of 0.66, 0.68, and 0.63, respectively, suggesting a small influence of the variables to the prediction. This suggests that the chosen rule of “chronic” PA for two

Prediction year = 2013		Outcome		
		Positive samples	No positive samples	Total
Prediction	Positive samples	1801 (97%)	60 (3%)	1861
	No positive samples	811 (44%)	1038 (56%)	1849
	Total	2612	1098	3710

TABLE 4 Contingency table comparing the definition of two previous years' chronic *Pseudomonas aeruginosa* (PA) status (2011 and 2012) to predict a patient having positive PA samples within the current year (2013) compared with a confirmed positive PA sample in that year (2013)

Variable	Model 1 (2013) Coefficient (SE)	Model 2 (2014)	Model 3 (2015)
Age	a*	a*	a
BMI	−0.013 (0.047)	−0.075 (0.039)	−0.033 (0.042)
CF-related diabetes	0.730 (0.373)	0.717 (0.337)*	−0.267 (0.283)
FEV ₁ (% pred)	−0.279 (0.737)	−0.888 (0.714)	−0.219 (0.705)
IV days	0.006 (0.005)	0.002 (0.005)	0.012 (0.006)
Pancreatic insufficiency	0.365 (0.446)	0.017 (0.426)	0.841 (0.390)*
Female sex	0.017 (0.296)	0.450 (0.275)	0.062 (0.270)
IMD	0.009 (0.010)	0.002 (0.009)	0.010 (0.009)
Area under ROC curve	0.664	0.678	0.628

TABLE 5 Regression model output using the outcome of a patient having positive PA samples within a given year and patient characteristics as predictors

Note: * $P < .05$; ** $P < .01$.

Abbreviations: BMI, body mass index; CF, cystic fibrosis; IMD, indices of multiple deprivation; IV, intravenous; PA, *Pseudomonas aeruginosa*.

^aAge was nonlinear with different functional forms: its effect on PPV plateaued after the age of 30 in all cases.

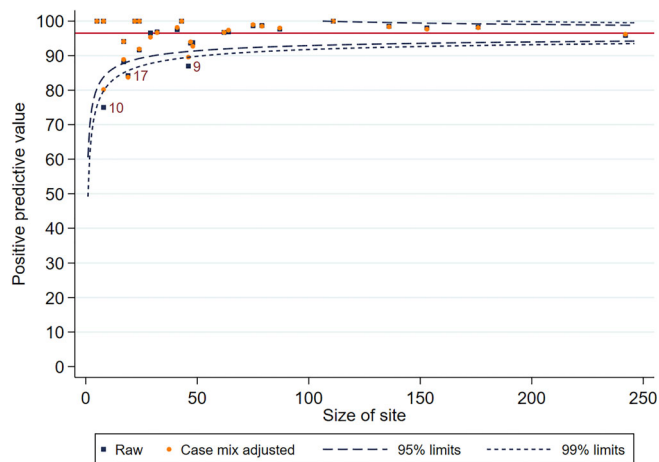


FIGURE 3 Funnel plot for adjusted and unadjusted PPV by center for 2013 predictions, lines reflect 95% (inner) and 99% (outer) reference intervals

successive years without an adjustment for patient characteristics is sufficient.

Furthermore, the adjusted PPV at center level is very similar to the unadjusted PPV, as shown in the funnel plot for 2015 (Figure 3). The difference in the precision of the prediction is minimal, giving a strong rationale that adjusting for covariates does not substantially affect the accuracy of the predictions.

4 | DISCUSSION

Patients with positive PA samples can be accurately identified in advance using routine data and a simple clinical rule. Patients with a chronic PA status reported within the UK CF Registry for the previous two consecutive years are likely to remain positive in the subsequent year, with minimal variation between patients of differing presenting characteristics or across treatment centers.

There are limitations to this work due to the data architecture of the UK CF Registry. Firstly, if PA status is reported as “none,” this could be because a person did not have PA infection, but alternatively may arise by the patient not producing any sputum at clinic visits or by missing data. Secondly, advances in CF treatments, including treatments for PA infection,^{18–20} means that the definition in this study should be re-explored again with future registry data. Thirdly, although guidance on the definition of chronic and intermittent *P. aeruginosa* status is provided by the CF Registry, it is uncertain how closely this definition is followed by individual CF centers. As encounter-based data are not available for the data CF Registry dataset that was used, confirmation was not possible, and so an important next step will be to validate this rule against a gold standard definition such as the Leeds criteria,²¹ a formally elicited consultant judgment or a consensus definition.

Nevertheless, this prediction rule appears robust to differing patient characteristics and may allow center comparisons to be carried out using registry data. This is important, since center comparisons underpin important advances in CF. For example, the comparisons of CF centers

in Toronto and Boston in the 1980s identifying the importance of aggressive nutritional support²² is now a standard of care in CF.²³

Due to the low prevalence of CF and the high numbers of patients required to make comparisons based on key outcomes such as FEV₁,²⁴ process measures such as adherence are important in providing sensitive measures to identify differences in quality of care²⁵ with higher adherence to treatments linked to better health outcomes.²⁶ The subgroup of people predicted to have PA in the current year based on historic registry data provide a homogenous group in whom to compare adherence rates with consensus about minimum treatment regimens (mucolytic plus antibiotic), allowing a minimum denominator to be imputed if prescription data are unavailable²⁷⁻²⁹).

Additionally, the use of convenience samples is a well-understood limitation of comparing adherence between cohorts or centers.³⁰ By using this prediction rule to identify patients from a registry dataset (more than 95% complete), researchers can understand to what extent reported adherence data in a center has included all patients with chronic PA. Identification of the number of patients included in adherence rates within a center, compared with the numbers in the historic registry data, allows an understanding of how many patients are missing, thus illuminating the potential impact of sampling error on estimates of center-level adherence. This work supports the national indicator for CF patients³¹ that has been developed to use adherence data from people with chronic PA to assess center-level adherence and establish the missing patients within the center.

5 | CONCLUSION

The routine use of data-logging nebulizers in CF allows the number of nebulizer doses used to be captured accurately at an individual level, but there is uncertainty regarding the method to robustly compare system-level adherence between different CF centers. The identification of a subgroup of adults predicted to have PA infection using routine data provides a first step toward destination robust comparison of system-level adherence between centers since there is consensus regarding treatment targets in this subgroup. We have developed a method of using routine data within the UK CF Registry to identify a subgroup of adults who are predicted, with good accuracy, to have a PA infection in the given year. The next step is to validate our method against other definitions of chronic PA infection.

ACKNOWLEDGEMENTS

We would like to thank the UK CF Trust for supplying the data that made this analysis possible. We would also like to acknowledge all the people with CF in the United Kingdom who consent to be part of the UK CF Registry.

FUNDING

This output is part of the CFHealthHub Data Observatory study, which is funded by NHS England Commissioning for Quality and Innovation.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

TRANSPARENCY STATEMENT

The authors confirm this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

AUTHORS' CONTRIBUTIONS

Conceptualization: Jen Lewis, Zhe Hui Hoo, Daniel Hind, Carla Girling, Martin Wildman, Elizabeth Shepherd, Julia Nightingale, Thomas Daniels, Jane Dewar, Sophie Dawson, Mary Carroll, Mark Allenby, Frank Edenborough, Rachael Curley, Charlotte Carolan.

Data Curation: Nikki Totton, Mike Bradburn, Jen Lewis, Zhe Hui Hoo, Carla Girling.

Formal Analysis: Nikki Totton, Mike Bradburn, Jen Lewis.

Funding Acquisition: Daniel Hind, Martin Wildman.

Methodology: Nikki Totton, Mike Bradburn, Jen Lewis, Zhe Hui Hoo, Martin Wildman.

Project Administration: Carla Girling.

Validation: Mike Bradburn.

Visualization: Nikki Totton, Mike Bradburn, Jen Lewis.

Writing—Original Draft Preparation: Nikki Totton, Zhe Hui Hoo, Daniel Hind, Martin Wildman.

Writing—Review and Editing: Zhe Hui Hoo, Mike Bradburn, Jen Lewis, Carla Girling, Elizabeth Shepherd, Julia Nightingale, Thomas Daniels, Jane Dewar, Sophie Dawson, Mary Carroll, Mark Allenby, Frank Edenborough, Rachael Curley, Charlotte Carolan.

Martin Wildman had full access to all of the data in the study and takes complete responsibility for the integrity of the data and the accuracy of the data analysis.

CONSENT TO PARTICIPATE

De-identified data on adults with cystic fibrosis was requested and obtained from the CF Registry in accordance with ethical approval.

DATA AVAILABILITY STATEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request subject to CF Registry approval.

ETHICS STATEMENT

National Health Service (NHS) research ethics approval was granted for the use of UK CF Registry data (Huntingdon Research Ethics Committee 07/Q0104/2) and under these terms, the UK CF Trust Steering Committee approved this study.

ORCID

Nikki Totton  <https://orcid.org/0000-0002-1900-2773>

Zhe Hui Hoo  <https://orcid.org/0000-0002-7067-3783>

REFERENCES

- Ryan G, Singh M, Dwan K. Inhaled antibiotics for long-term therapy in cystic fibrosis. *Cochrane Database Syst Rev*. 2011;3:CD001021. <https://doi.org/10.1002/14651858.CD001021.pub2>
- Quittner AL, Zhang J, Marynchenko M, et al. Pulmonary medication adherence and health-care use in cystic fibrosis. *Chest*. 2014;146:142-151. <https://doi.org/10.1378/chest.13-1926>
- Daniels T, Goodacre L, Sutton C, Pollard K, Conway S, Peckham D. Accurate assessment of adherence. *Chest*. 2011;140(2):425-432. <https://doi.org/10.1378/chest.09-3074>
- Quinton HB, O'Connor GT. Current issues in quality improvement in cystic fibrosis. *Clin Chest Med*. 2007;28:459-472. <https://doi.org/10.1016/j.ccm.2007.02.013>
- Geller DE, Madge S. Technological and behavioral strategies to reduce treatment burden and improve adherence to inhaled antibiotics in cystic fibrosis. *Respir Med*. 2011;105:S24-S31. [https://doi.org/10.1016/S0954-6111\(11\)70024-5](https://doi.org/10.1016/S0954-6111(11)70024-5)
- CFHH Collaborative. CFHealthHub. 2019. <https://www.cfhealthhub.com/>. Accessed July 30, 2021.
- Emerson J, Rosenfeld M, McNamara S, Ramsey B, Gibson RL. Pseudomonas aeruginosa and other predictors of mortality and morbidity in young children with cystic fibrosis. *Pediatr Pulmonol*. 2002;34:91-100. <https://doi.org/10.1002/ppul.10127>
- Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J*. 2012;40(6):1324-1343. <https://doi.org/10.1183/09031936.00080312>
- Abel G, Payne R, Barclay M. UK Deprivation Indices. 2016. <https://doi.org/10.5523/bris.1ef3q32gybk001v77c1ifmty7x>. Accessed July 30, 2021.
- The UK. Cystic Fibrosis Trust Diabetes Working Group. Management of cystic fibrosis related diabetes mellitus. 2004. <https://www.cysticfibrosis.org.uk/sites/default/files/2020-12/Diabetes%20mellitus%20management%20Jun%2004.pdf>. Accessed July 30, 2021.
- Dickson RP, Erb-Downward JR, Freeman CM, et al. Changes in the lung microbiome following lung transplantation include the emergence of two distinct pseudomonas species with distinct clinical associations. *PLoS One*. 2014;9(5):e97214. <https://doi.org/10.1371/journal.pone.0097214>
- Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc*. 1979;74:829-836. <https://doi.org/10.1080/01621459.1979.10481038>
- Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Stat*. 1994;43:429. <https://doi.org/10.2307/2986270>
- Zemanick ET, Wagner BD, Robertson CE, et al. Airway microbiota across age and disease spectrum in cystic fibrosis. *Eur Respir J*. 2017;50(5):1700832. <https://doi.org/10.1183/13993003.00832-2017>
- Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med*. 2005;24(8):1185-1202. <https://doi.org/10.1002/sim.1970>
- R Core Team. R: A Language and Environment for Statistical Computing. R Foundational Statistical Computing. Vienna, Austria. 2017. www.R-project.org/
- StataCorp. Stata Statistical Software: Release 15, 2017.
- Ren E, Zhang C, Li D, Pang X, Liu G. Leveraging metal oxide nanoparticles for bacteria tracing and eradicating. *View*. 2020;1(3):20200052. <https://doi.org/10.1002/viw.20200052>
- Hu D, Zou L, Gao Y, Jin Q, Ji J. Emerging nanobiomaterials against bacterial infections in postantibiotic era. *View*. 2020;1(3):20200014. <https://doi.org/10.1002/viw.20200014>
- Xiu W, Shan J, Yang K, Xiao H, Yuwen L, Wang L. Recent development of nanomedicine for the treatment of bacterial biofilm infections. *View*. 2021;2(1):20200065. <https://doi.org/10.1002/viw.20200065>
- Lee TWR, Brownlee KG, Conway SP, Denton M, Littlewood JM. Evaluation of a new definition for chronic Pseudomonas aeruginosa infection in cystic fibrosis patients. *J Cyst Fibros*. 2003;2:29-34. [https://doi.org/10.1016/S1569-1993\(02\)00141-8](https://doi.org/10.1016/S1569-1993(02)00141-8)
- Corey M, McLaughlin FJ, Williams M, Levison H. A comparison of survival, growth, and pulmonary function in patients with cystic fibrosis in Boston and Toronto. *J Clin Epidemiol*. 1988;41:583-591. [https://doi.org/10.1016/0895-4356\(88\)90063-7](https://doi.org/10.1016/0895-4356(88)90063-7)
- Stallings VA, Stark LJ, Robinson KA, Feranchak AP, Quinton H. Evidence-based practice recommendations for nutrition-related management of children and adults with cystic fibrosis and pancreatic insufficiency: results of a systematic review. *J Am Diet Assoc*. 2008;108:832-839. <https://doi.org/10.1016/j.jada.2008.02.020>
- Nightingale JA, Osmond C. Does current reporting of lung function by the UK cystic fibrosis registry allow a fair comparison of adult centres? *J Cyst Fibros*. 2017;16:585-591. <https://doi.org/10.1016/j.jcf.2017.04.007>
- Mant J. Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care*. 2001;13:475-480. <https://doi.org/10.1093/intqhc/13.6.475>
- Eakin MN, Bilderback A, Boyle MP, Mogayzel PJ, Riekert KA. Longitudinal association between medication adherence and lung health in people with cystic fibrosis. *J Cyst Fibros*. 2011;10(4):258-264. <https://doi.org/10.1016/j.jcf.2011.03.005>
- UK Cystic Fibrosis Trust Antibiotic Working Group. Antibiotic treatment for cystic fibrosis, 2009. <https://www.cysticfibrosis.org.uk/sites/default/files/2020-11/Anitbiotic%20Treatment.pdf>. Accessed July 30, 2021.
- Smyth AR, Bell SC, Bojcin S, et al. European cystic fibrosis society standards of care: best practice guidelines. *J Cyst Fibros*. 2014;13:S23-S42. <https://doi.org/10.1016/j.jcf.2014.03.010>
- Mogayzel PJ, Naureckas ET, Robinson KA, et al. Cystic fibrosis pulmonary guidelines: chronic medications for maintenance of lung health. *Am J Respir Crit Care Med*. 2013;187(7):680-689. <https://doi.org/10.1164/rccm.201207-1160OE>
- Bender B. Adherence intervention research what have we learned and what do we do next? *J Allergy Clin Immunol*. 2003;112(3):489-494. [https://doi.org/10.1016/S0091-6749\(03\)01718-4](https://doi.org/10.1016/S0091-6749(03)01718-4)
- Wildman M. Normative adherence to nebulised therapy in cystic fibrosis for patients with chronic pseudomonas acquisition, 2020. <https://www.nice.org.uk/standards-and-indicators/nlindicators/normative-adherence-to-nebulised-therapy-in-cystic-fibrosis-for-patients-with-chronic-pseudomonas-acquisition>. Accessed July 30, 2021.

How to cite this article: Totton N, Bradburn M, Hoo ZH, et al. Prospectively predicting *Pseudomonas aeruginosa* infection/s using routine data from the UK cystic fibrosis register. *Health Sci Rep*. 2021;4:e381. doi:10.1002/hsr2.381

APPENDIX

TABLE A1 Contingency table comparing the definition of two previous years' chronic PA status to predict a patient having positive PA samples within the current year (2014 and 2015 respectively) compared with a confirmed positive PA sample in that year (2014 and 2015 respectively)

Prediction year = 2014		Outcome		
		Positive samples	No positive samples	Total
Prediction	Positive samples	1804 (96%)	68 (4%)	1872
	No positive samples	826 (41%)	1189 (59%)	2015
Total		2630	1257	3887

Prediction year = 2015		Outcome		
		Positive samples	No positive samples	Total
Prediction	Positive samples	1782 (96%)	70 (4%)	1852
	No positive samples	859 (39%)	1333 (61%)	2192
Total		2641	1403	4044

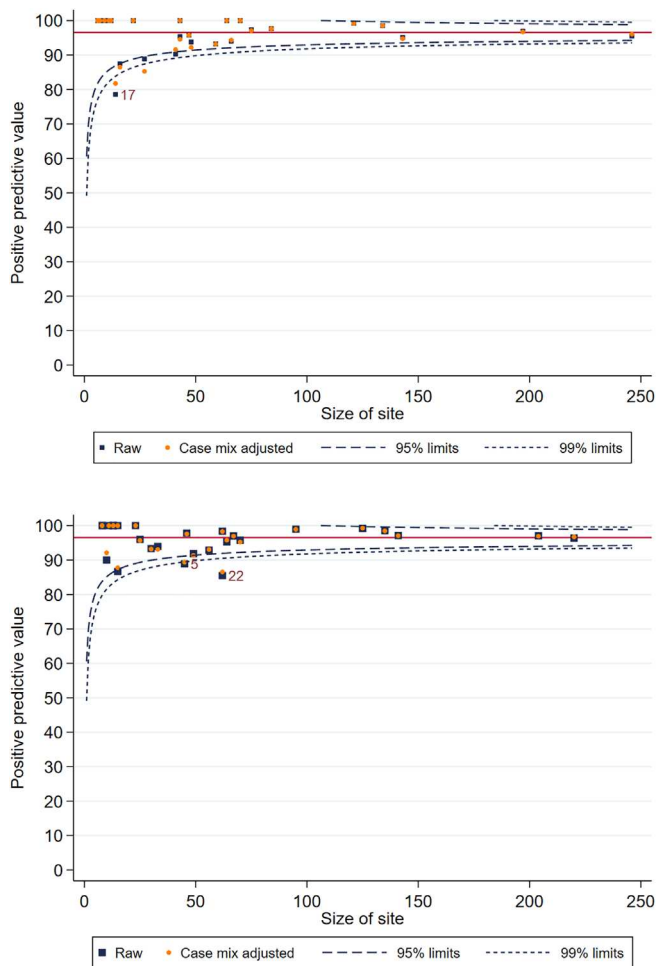


FIGURE A1 Funnel plot for adjusted and unadjusted PPV by center for 2014 (top) and 2015 (bottom) predictions, lines reflect 95% (inner) and 99% (outer) reference intervals

TABLE A2 Summary of CF center-level PPV, the number and proportion of patients that are predicted to have positive PA samples in the given year

	Prediction year		
	2013	2014	2015
PPV			
Median	98%	97%	97%
Min	85%	86%	86%
Max	100%	100%	100%
IQR	94%-99%	94%-100%	94%-99%
Number predicted to have positive PA samples			
Median	49	53	52
Min	8	10	9
Max	273	279	249
IQR	27-82.5	22-82.5	19.5-76
Proportion predicted to have positive PA samples			
Median	40%	37%	37%
Min	13%	8%	10%
Max	53%	55%	55%
IQR	32%-46%	31%-46%	28%-44%

Abbreviations: CF, cystic fibrosis; PA, *Pseudomonas aeruginosa*; PPV, positive predictive value.