

Development and temporal evaluation of a multivariable prediction model to screen for late-diagnosed cystic fibrosis

Richard Belkin

rbelkin@sblung.com

Santa Barbara Cottage Hospital

Yuri Matusov

Santa Barbara Cottage Hospital

Barry Belkin

Research Article

Keywords: Screening, Algorithm, Prediction, Adult Cystic Fibrosis

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11176895/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

1 **Long Title**

2 **Development and temporal evaluation of a multivariable prediction model to screen for**

3 **late-diagnosed cystic fibrosis**

4 **Short Title**

5 Prediction model for cystic fibrosis

6 **Author and Affiliation List**

7 Richard A Belkin, MD, MSCE, FACP, FCCP^{1#a*}

8 Yuri Matusov, MD, FACP, ATSF²

9 Barry Belkin, PhD³

10 ¹Department of Pulmonary and Critical Care Medicine, Santa Barbara Cottage Hospital, Santa Barbara, CA,
11 United States of America

12 ²Division of Pulmonary and Critical Care Medicine, Department of Medicine, Cedars-Sinai Medical Center,
13 Los Angeles, CA, United States of America

14 ³Independent mathematician/statistician, Santa Barbara, CA, United States of America

15 ^{#a}Current Address 2403 Castillo Street, Suite #206, Santa Barbara, CA 93105

16 *Corresponding author

17 rbelkin@sblung.com (RB)

18

19 Abstract

20 **Background.** Adult CF diagnosis is challenging because symptoms are often atypical. We developed and
21 tested a screening model to identify patients requiring diagnostic testing.

22 **Methods.** Adults evaluated for possible CF at one accredited CF center underwent sweat chloride testing for
23 features suggesting CF or CFTR-related disease. Dataset 1 (171 patients, 2005–2021) supported model
24 development and internal testing; Dataset 2 (85 patients, 2021–2023) provided temporal evaluation. Under CF
25 Foundation guidelines, Dataset 1 included 25 CF-likely (Type A), 74 CF-not-resolved (Type B), and 72 CF-
26 unlikely (Type C) patients; Dataset 2 included 4, 27, and 54, respectively. The multivariable probabilistic
27 model used seven observable CF-related conditions and recommended testing when either: (1) a logistic-
28 regression patient score S_A met ROC-derived threshold C_0 ; or (2) bronchiectasis/tree-in-bud was present with
29 negative or unknown findings for the other six conditions.

30 **Results.** Predictors were pancreatic insufficiency, acute or chronic pancreatitis, Gram-negative infection,
31 Gram-positive infection, NTM infection, bronchiectasis/tree-in-bud, and family history of a CFTR gene
32 variant. With either criterion, sensitivity for recommending testing was 1.00 for Types A and B in both
33 datasets; among Type C patients considered false positives by clinical judgment, 2/126 (1.59%) were model
34 true negatives. With Criterion 1 alone, Type A sensitivity was 1.00 in both datasets. Type B1 denoted one CF
35 gene mutation and Type B0 otherwise; sensitivity was below 1.00 for both subgroups in both datasets.
36 Bootstrap 95% CIs were: Dataset 1—Type B0 sensitivity [.675,1.00], Type B1 sensitivity [.903,.975], Type C
37 specificity [.058,.219], and AUC [.703,.929]; Dataset 2—Type B0 sensitivity [.730,.951], Type B1 sensitivity
38 [.885,1.00], Type C specificity [.040,.391], and AUC [.735,.913]. For the two-criterion rule, bootstrapping was
39 unnecessary for Type A, B0, and B1 sensitivity because each equaled 1.00 in both datasets; too few Type C
40 true negatives supported useful dataset-specific bootstrap specificity intervals.

41 **Conclusions.** Applied separately, the model recommended testing for every Type A and B patient but did not
42 recommend testing for only a few Type C patients, closely matching clinical judgment. Findings are
43 hypothesis-generating and require prospective multicenter temporal evaluation.

44 **Trial registration.** Not applicable.

45 **Key Words List**

46 Screening; Algorithm; Prediction; Adult Cystic Fibrosis

47

48 **1. Introduction**

49 Despite newborn screening, there is a clinically significant proportion of both CF patients and CF-RD patients
50 whose diagnosis is delayed into adulthood or may even go undiagnosed entirely(1). Screening for CF in adults
51 is currently based on society guideline–endorsed criteria(2,3). Late-diagnosed CF is highly variable,
52 heterogeneous with an atypical presentation, and needs to be considered for all patient ages, races, and
53 ethnicities. The adult patient symptoms and findings are not specific to CF and may be so insidious that a
54 diagnosis of CF or CFTR-RD is not considered even by an experienced clinician.(4–7) Health inequities in CF
55 diagnosis may arise because access to specialized CF evaluation and diagnostic testing varies across
56 geographic and sociodemographic groups, potentially contributing to delayed or missed diagnosis in
57 underserved populations. With the advent of CFTR modulator therapy, increasing numbers of adults with
58 bronchiectasis coming to medical attention in the limited centers with concentrated expertise in CF/non-CF
59 bronchiectasis, and rising healthcare costs, it is important to develop a framework to assist the clinician in the
60 determination of which patients should be referred for CF genetic and sweat chloride testing.(2,8) The most
61 significant benefits of an accurate diagnosis of CF are access to highly effective CFTR modulator therapy
62 (9,10), other emerging CF genetic therapies (11) and multidisciplinary symptomatic care.(12) Establishing a
63 definitive diagnosis of CF in adulthood facilitates genetic counseling for patients and their family members,
64 including reproductive risk assessment and carrier testing, while potentially resolving longstanding diagnostic
65 uncertainty.(13,14) It is also important to minimize CF testing when CF is unlikely in order to decrease cost,
66 time/burden, and patient anxiety that may develop when CF testing is ordered.

67 To address the multiple challenges of adult CF diagnosis, an active screening program for sweat chloride
68 testing was initiated at our CF center in 2005 through outreach and education of our medical community.

69 The present study had two primary objectives: The first objective was to develop a multivariable probabilistic
70 screening model whose function is to recommend whether a patient should or should not undergo diagnostic
71 laboratory testing for CF (sweat chloride testing, and when indicated, subsequent CFTR genetic testing). The

72 second objective was to test the predictive accuracy of the screening model when separately applied to each
73 of two datasets.

74

75 **2. Materials and Methods**

76 The study was approved by the Santa Barbara Cottage Hospital Institutional Review Board (IRB #16-94) on
77 2/15/2017. The IRB approved waivers of consent and HIPAA authorization.

78 **Patient and Public Involvement:** Patients and/or the public were not directly involved in the design or
79 implementation of this study.

80 **2.1 Study type, sources of data, participants, and outcome.** This is a single-center, retrospective,
81 observational study of adult patients aged ≥ 18 years old who were: (i) screened for CF by affiliated physicians,
82 predominantly pulmonologists and gastroenterologists, using a sweat chloride test at Santa Barbara Cottage
83 Hospital (SBCH), or (ii) referred to and evaluated at the Cottage Cystic Fibrosis and Bronchiectasis Center,
84 with a diagnosis of CF under consideration and with at least one completed sweat chloride test. The Cottage
85 Cystic Fibrosis and Bronchiectasis Center is a CFF-accredited, multidisciplinary CF program affiliated within
86 a community teaching hospital and functions as a tertiary referral center both for patients with newly diagnosed
87 adult CF and patients for whom the diagnosis is being considered. The patients included in the study were
88 screened for CF via sweat chloride testing between January 2005 and December 2023. The age listed for each
89 patient was based on the date that the first sweat test was performed. Race/ethnicity were determined by chart
90 review. The same data preprocessing and quality-checking procedures were applied to all participants, without
91 differences based on sociodemographic characteristics. Prior to analysis, the datasets were reviewed for
92 completeness, consistency, duplicate records, and implausible or out-of-range values. Predictor variables and
93 outcome classifications were checked against the predefined study definitions, and discrepancies were resolved
94 by author RB through review of the source clinical record. Of 258 potentially eligible individuals between
95 both datasets, 2 were excluded because of insufficient clinical data, leaving a total of 256 included patients.

96 The screening model is based on nonrandom split-sample development and testing methods consistent with
97 TRIPOD+AI guidelines.(15) The initial study population (Dataset 1) was comprised of 171 patients tested for
98 CF between January 2005 and April 2021. This dataset was used both in the development and internal testing
99 of the screening model. A second study population (Dataset 2) was comprised of 85 patients tested for CF
100 between May 2021 and December 2023 and was used in the temporal evaluation of the screening model.
101 Patients were excluded from the study if insufficient clinical data were available. The development and
102 evaluation cohorts were derived from the same healthcare setting and used the same eligibility criteria,
103 outcome definitions, and clinical predictor definitions. The cohorts differed only by study period, as stated
104 above. The specific function of the screening model is to determine whether or not the available clinical
105 evidence supports a recommendation that the patient should be tested for CF.

106 **2.2 Methods Overview:** The basic analysis framework for the screening model is described below. Detailed
107 explanatory material, including the supporting mathematical and probabilistic methods, is provided in
108 Additional file 1.

109 Based on CF testing as outlined in Figure 1 and after thorough clinical chart review, patients were classified
110 as follows: Type *A* (CF likely), Type *B* (CF not resolved (CFNR), including CFTR-RD), or Type *C* (CF
111 unlikely). This CF classification scheme is based on the CFF diagnosis guidelines (16–18) specified in Figure
112 1. Outcome classification was performed by the medical director of the Cystic Fibrosis Center (RB), a board-
113 certified pulmonologist with extensive experience in the diagnosis and management of CF and CFTR-related
114 disorders. Demographic characteristics of the outcome assessor were not considered relevant to the clinical
115 classification process and were not collected for this study. Classification was based on review of the clinical
116 record and Cystic Fibrosis Foundation diagnostic guidelines (16) and was performed independently of model
117 development. At the time of classification, RB was unaware of the subsequently derived prediction model,
118 model coefficients, and the calculated prediction scores. Predictor data were abstracted from the medical record
119 by the study investigators, including internal medicine residents and trained research assistants. Demographic

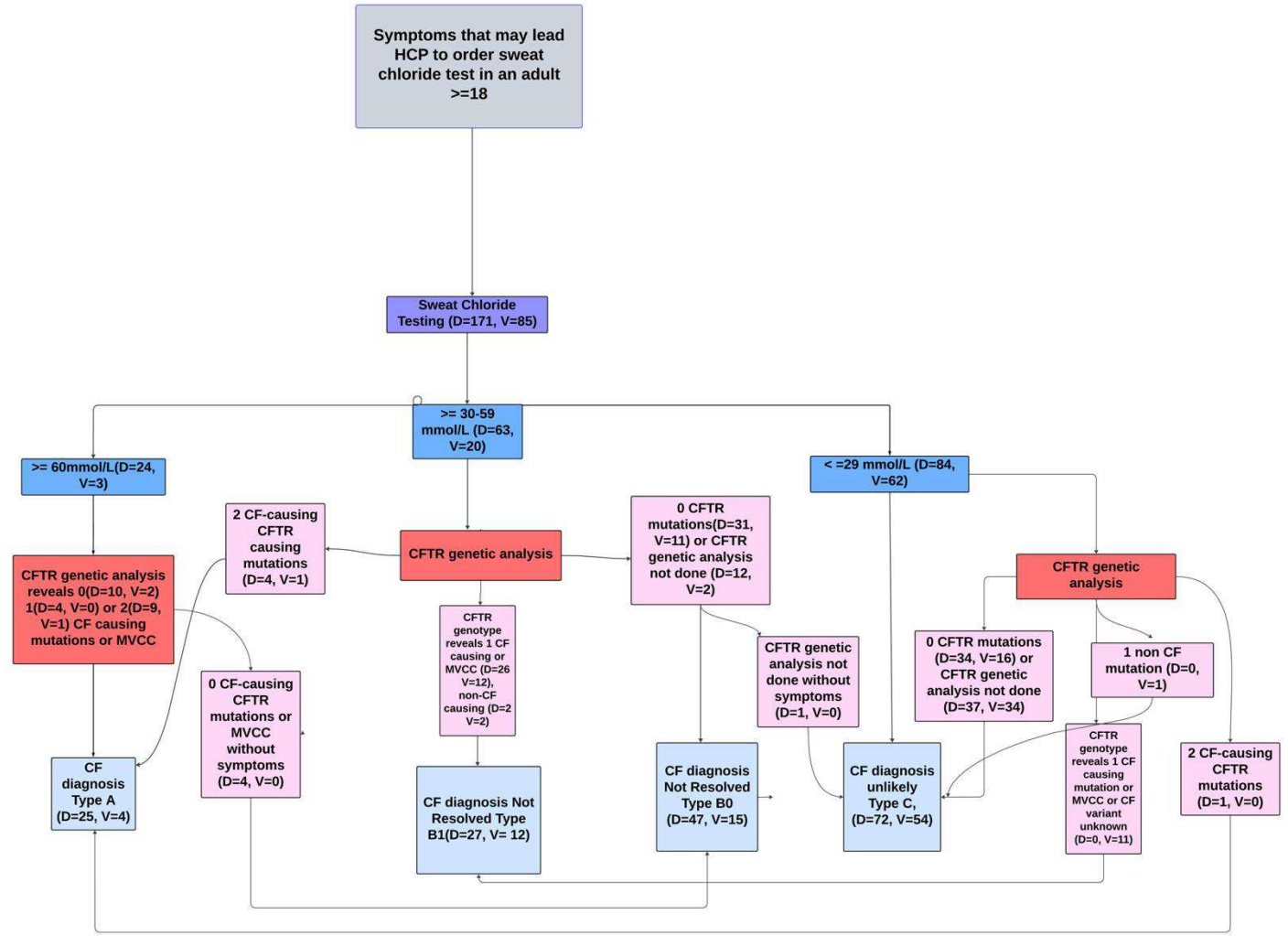
120 characteristics of the predictor-data abstractors were not collected because they were not considered relevant
121 to the objective abstraction of predefined clinical variables.

122

123 **Figure 1:** Flowchart for diagnosis of CF and CFNR (includes patients with CFTR-RD). D = Development
124 and Internal Testing Cohort (Dataset 1) and V=Temporal evaluation Cohort (Dataset 2).

125

126



127 **Figure 1 Legend:** Diagnosis of CF based on CF Foundation diagnosis guidelines.(16,17) Of note, our CF
128 center does not have routine access to nasal potential difference (NPD) or intestinal current measurement
129 testing to assist in CF diagnoses.

Type *B* was partitioned into two subgroups: (1) patients (designated Type *B*₁) with one CF causing CFTR gene variant or VVCC, and (2) patients (designated Type *B*₀) who were either (a) reported with no CFTR gene variant, (b) reported with a non-CF-causing variant, or (c) not tested for CFTR gene variants. See Table S3-D and S3-V for breakdown of Type *B* patients. The importance of testing and identifying Type *B*₁ patients for CF lies in the potential candidacy of a Type *B*₁ patient for CFTR modulator or other genetic therapies for CF. However, the intended performance goal for the screening model is to recommend for CF testing all Type *A*, Type *B*₁, and Type *B*₀ patients. *A*, *B*₁, and *B*₀ patients recommended for CF testing by the prediction model are counted as true positives. Type *C* patients recommended for CF testing by the screening model are counted as false positives. See the attached diagnosis flowchart, Figure 1.

2.3 Clinical predictors. The study data included 23 potential LR model variables (Supplement Table S4-D and S4-V) referred to as *clinical predictors*, each of which records the presence (value=1), absence (value=0), or the unknown diagnosis of a specific clinical condition (also value=0) known to be observed in Type *A* patients. Candidate clinical predictors were chosen based on prior knowledge of their presence in adult diagnosed CF patients and on their potential to be predictive of CF. Patient data for the particular clinical predictors selected for possible use in screening for CF testing using predefined study definitions was extracted from retrospective electronic chart review by the authors (RB and YM) and research assistants using EPIC (Epic Systems, Verona, WI) and Elation (Elation Health, San Francisco, CA). These data were managed using RED Cap electronic data capture tools hosted at SBCH (19,20) and MS Excel (Version 16.0.14931.20132, Redmond, WA). Ambiguous or incomplete entries were resolved through review of the source medical record.

A given clinical condition was determined to be present in a patient if it was present on chart review at the time data collection was performed. BMI and FEV1 were collected at the time point closest to when sweat testing was performed. Pancreatic insufficiency was determined to be present if listed in the patient's clinical chart and/or if the patient had a low pancreatic fecal elastase level (<500 µg/g). The presence of bronchiectasis and tree-in-bud opacities was identified either in a chest CT report or in the problem

list/assessment, if CT reports were not available. Other variables such as acute or chronic pancreatitis were determined to be present based on detailed review of the patient’s medical chart. See the supplementary material (Additional File 1) for the list of clinical conditions collected and considered in the prediction model.

Clinical relevance and importance testing based on statistical metrics was used to reduce the candidate clinical conditions under consideration to the subset of 15 clinical conditions used in the screening model. One of the two metrics, the prevalence ratio (PR), measures the degree to which knowledge that a patient is diagnosed as positive for a specified CF-related clinical condition increases the prevalence of the CF Type A classification. For a specified CF-related clinical condition, the PR value is given by the following ratio:

$$\frac{\text{Fraction of Dataset 1 patients positive for the CF-related clinical condition who are Type } A}{\text{Fraction of Dataset 1 patients who are Type } A}$$

The second metric measures the prevalence of a specified CF-related clinical condition in patients classified as Type A.

In some cases, composites of multiple candidate clinical conditions were formed, such that a patient is considered to be positive for the composite condition if positive for one or more of the associated component conditions. This reduced the number of individual or composite clinical conditions used in the screening model from 15 to the 7 listed in Table 1.

2.4 Sample size: No formal a priori sample-size calculation was performed. The sample size was determined by the number of eligible patients available during the predefined study period. The available sample size and number of outcome events constrained model complexity and the precision of estimated model performance, particularly in the temporal evaluation cohort, which included only four Type A patients.

2.5 Statistical Analysis. The screening model decision rule recommends a patient for CF testing if either one or both of the following requirements are satisfied: (Criterion 1) a calculated numerical patient score S_A with a value between 0 and 1 equals or exceeds a threshold value C_0 determined using logistic regression (LR)

180 and ROC curve analysis; (Criterion 2) the patient diagnosis is positive for bronchiectasis/tree-in-bud and
181 either negative or unknown for the remaining 6 CF-related clinical conditions.

182 The value of S_A for a given patient quantifies the weight of clinical evidence supporting a Type A
183 classification and is determined using LR analysis in the manner described in Section S.1 of the Supplement.
184 As reported in the Results Section below, for all Dataset 1 and Dataset 2 patients classified as Type A
185 patients, Criterion 1 is satisfied. As a result, all Type A patients in the two datasets are recommended for CF
186 testing by the screening model. However, it was observed that it is not the case for Dataset 1 patients that
187 Criterion 1 is satisfied for all Type B patients. The LR analysis results for Dataset 1 indicated that all Type B
188 patients not recommended for CF testing based on Criterion (1) also have bronchiectasis/tree-in-bud present
189 and either a negative or unknown diagnosis for each of the 6 remaining CF-related clinical conditions. Thus,
190 criterion (2) is met for these patients, and all Dataset 1 Type B patients are recommended for CF testing by
191 the screening model. On that basis, criterion 2 was incorporated into the screening model decision rule for
192 CF testing. It was subsequently observed that with both criterion 1 and criterion 2 incorporated into the
193 decision rule for CF testing, all Dataset 2 Type B patients are recommended for CF testing by the screening
194 model. It was also observed to be the case that criterion (2) is satisfied by some Type C patients not
195 recommended for CF testing based on Criterion 1.

196 **2.6 Use of multiple LR models.** Of the 7 CF-related clinical conditions (listed in Table 1) selected for use in
197 the LR analysis, the average number of conditions for which a Dataset 1 patient is positive is 2.01. Of the 171
198 Dataset 1 patients, only 16 patients are positive for more than 3 of the selected clinical conditions. The set of
199 7 selected clinical conditions has 35 subsets of size 3. The approach taken then is that LR analysis with 3
200 clinical predictors as the variables is separately applied to each of these 35 subsets of size 3. For each patient,
201 each of the 35 LR models produces an estimate for the probability that the patient CF classification is Type A .
202 The value of the score S_A for the patient is taken to be the maximum of these 35 probability estimates. The
203 decision based on LR analysis whether or not to recommend a patient for CF testing is thus based on the
204 particular set of 3 clinical predictors that provides the strongest clinical evidence that the patient CF

classification is Type A. The choice of the specific CF-related clinical conditions used in the LR analysis is based on the following metrics: (1) for each of the candidate CF-related abnormalities, a measure of the degree to which for Dataset 1 patients, positivity for that abnormality increases the likelihood that the patient CF classification is Type A; (2) for Dataset 1 patients, the prevalence of each candidate CF-related abnormality; (3) biologic plausibility and ease of availability to the clinician.

Table 1. List of 7 CF-Related Clinical Conditions or Composite Clinical Conditions Selected for Use in LR Analysis

	CF-related clinical condition or composite clinical condition	Definition/components
1	Pancreatic insufficiency	Pancreatic insufficiency documented clinically or by low fecal elastase.
2	Acute and/or chronic pancreatitis	Acute, recurrent, and/or chronic pancreatitis.
3	Gram-negative bacterial infection	<i>Pseudomonas aeruginosa</i> , <i>Alcaligenes xylosoxidans</i> , <i>Haemophilus influenzae</i> , <i>Burkholderia cepacia</i> , <i>Stenotrophomonas</i> , and/or <i>Moraxella catarrhalis</i> .
4	Gram-positive bacterial infection	Methicillin-resistant <i>Staphylococcus aureus</i> (MRSA) and/or methicillin-sensitive <i>Staphylococcus aureus</i> (MSSA).
5	Nontuberculous mycobacterial infection	<i>Mycobacterium avium</i> complex (MAC) or non-MAC nontuberculous mycobacteria.
6	Bronchiectasis and/or tree-in-bud opacities	Bronchiectasis and/or tree-in-bud opacities identified on chest CT report or clinical assessment/problem list when CT report was unavailable.
7	Family history of CFTR gene variant	Family history of a CFTR gene variant.

Legend: List of CF-related clinical Conditions or Composite Clinical Conditions selected to be used based on clinical relevance and statistical metrics in logistic regression analysis

2.7 Estimation of LR model parameters. The values of the parameters for each of the 35 different 3-variate LR models are determined from Dataset 1 as described in Section S.2 of the Supplement. Each one of these 35 tables is linked to a particular set of 3 CF-related clinical predictors and records the estimated coefficients for each LR model having as its predictors a subset of that set of 3 predictors.

An important feature of the LR model parameter estimation is that the data for all 171 Dataset 1 patients are used in estimating the parameters of each of the 35 separate 3-variate LR models. This multiple-model approach was chosen to reduce the risk of unstable parameter estimation and potential model overfitting if for

each patient, S_A were based on a single 7-variate LR model. The lower-dimensional models were chosen to reduce parameter instability in this relatively small development dataset. Since the multi-model selection procedure may itself introduce optimism, evaluation using independent data was required.

2.8 LR decision rule for recommendation of CF testing. The value of S_A is compared with a threshold C_0 whose value is determined using ROC curve analysis. The decision rule used by the screening algorithm is the following: if $S_A \geq C_0$, then CF testing is recommended; if $S_A < C_0$, then CF testing is not recommended.

Choice of value for C_0 . An important subject of study investigation is whether the value of C_0 can be chosen so that if the screening model is separately applied to Datasets 1 and 2, in each case, two important conditions are met: (1) the calculated sensitivity value is 1 for recommending Type A patients for CF testing and (2) the calculated sensitivity value is 1 for recommending Type B patients for CF testing. The rationale for choosing 0.176 as the threshold value c_0 is that based on the Dataset 1 ROC curve, the false positive rate for patients classified as CF Type C is then minimized subject to the constraint that each Dataset 1 patient classified as CF Type A is a true positive.

2.9 Missing data. Patients for whom most of the clinical information was missing were excluded from the study. This applied to a very small number of patients. If the diagnosis for a specific clinical condition was missing for an individual patient, the diagnosis for that clinical condition was identified as unknown in the LR models used for development and temporal evaluation. Missing data were identified in one of three ways: testing not done due to low clinical evidence (MNAR), testing not ordered or completed (MAR), or test done with result truly missing (MAR). Most of the missing data were classified as MNAR. (See the Supplement for a breakdown of missing data in Datasets 1 and 2.) The number of predictors used in applying an LR model to a patient is adjusted to be equal to the number of CF-related clinical conditions for which there is diagnostic data. For example, in cases where the value of one of the predictor variables in a 3-variate LR model was truly unknown for a given patient, a bivariate LR model was substituted for 3-variate LR. See the Supplement for

247 frequency of missing data for each clinical condition in both datasets, and for further explanation of how
248 missing data was addressed.

249 **2.10 Predictive Accuracy of Screening Model.** The results of separately applying the screening model to
250 each dataset are in the form of three metrics: sensitivity values for Type A and Type B, and a specificity value
251 for Type C. The bootstrap method of randomized resampling was chosen as a means of quantifying the
252 estimation errors associated with these metrics. Bootstrapping: randomized resampling with replacement was
253 applied 200 times to the LR results for the 171 patient Dataset 1, and was used to generate a sampling
254 distribution for each of the following quantities: the AUC value for the ROC curve, the Type A sensitivity
255 value, the Type B1 sensitivity value, the Type B0 sensitivity value, and the Type C specificity value. For each
256 of these five values, the 95% confidence interval is determined from the corresponding sampling distribution.
257 A similar procedure was used to generate sampling distributions and 95% confidence intervals for Dataset 2.
258 The complete model calculations, ROC analyses, and Bootstrap results for Dataset 1 and Dataset 2 are
259 tabulated in Additional files 3 and 4, respectively. The testing of the predictive accuracy of the screening model
260 is further described in Section S.3 of Additional file 1.

261 **2.11 Illustrative example of application of screening model.** The application of the screening model to a
262 patient is described in detail in Section S.5 of the Supplement.

263 **2.12 Class imbalance.** No specific methods were used to address class imbalance; all eligible patients were
264 included without over- or undersampling or class weighting.

265 **2.13 Fairness.** No formal fairness analysis was performed. Because the study population lacked racial and
266 ethnic diversity, model performance across demographic subgroups could not be reliably assessed and should
267 be evaluated in future studies with larger, more diverse populations.

268 **2.14 Protocol and Registration.** A formal prespecified study protocol was not published; study procedures
269 were conducted under the protocol approved by the Santa Barbara Cottage Hospital Institutional Review Board
270 (IRB #16-94). This retrospective study was not registered.

271 **2.15 Data sharing:** The de-identified patient-level data for the development and temporal evaluation cohorts
272 are provided in Additional file 2. Direct patient identifiers have been removed to protect patient confidentiality.

273 **2.16 Code sharing:** No separate programming code was developed for this study. The two Excel workbooks
274 containing the model calculations, formulas, and statistical analyses are available in the accompanying
275 Additional files 3 and 4.

276 **2.17 Use of generative intelligence:** ChatGPT (OpenAI) was used for language editing, manuscript
277 organization, reporting-checklist review, and consistency checking. All AI-assisted output was reviewed and
278 verified by the authors, who take full responsibility for the manuscript's content.

279 **3. Results**

280 **3.1 Participants.** The baseline demographics and clinical characteristics can be found in Tables S1-5D and
281 S1-5V. In the 171-patient Dataset 1 cohort, 25 patients (15%) were classified as having CF, 74 (43%) were
282 classified as CFNR, and 72 (42%) were classified as CF or CF-related disease unlikely. These patients had a
283 median age of 53, 61, and 58 years, respectively, and all three groups had female predominance (70% of the
284 development cohort). Of the CF group, 20 (80%) had sweat chloride > 60 mmol/L, whereas 58 (78%) of the
285 CFNR group had sweat chloride between 30-59 mmol/L (refer to Table S1-D). In the 85-patient Dataset 2
286 cohort, 4 (5%) were classified as having CF, 27 (32%) were classified as CFNR, and 54 (64%) were classified
287 as CF unlikely. These patients had a median age of 59 (Dataset 1) and 62 (Dataset 2) years and had a female
288 predominance with a similar distribution of sweat chloride levels (refer to Table S1-V). Of note, the
289 development and evaluation cohorts were similar in age and sex distribution but differed in outcome
290 prevalence. CF was more prevalent in the development cohort (15% vs 5%), whereas the CF-unlikely
291 classification was more common in the evaluation cohort (64% vs 42%).

292 **3.2 Selection of CF-related clinical conditions and model development with Dataset 1.** As previously
293 described, the screening model is based on multiple LR models, each with at most 3 predictive variates. As the
294 number of CF-related clinical conditions considered in the LR analysis increases, so does the number of

different 3-variate LR models. With 7 CF-related clinical conditions, the number of different 3-variate LR models is 35 (the number of 3 element subsets of a set with 7 elements).

As previously indicated, 7 CF-related clinical conditions were chosen as the predictive variates (Table 2). Examples of the Type A prevalence ratio if the patient is positive for a particular clinical condition are: pancreatic insufficiency (PR =4.56), acute/chronic pancreatitis (PR=2.49), a family history of CFTR gene variants (PR=2.41), and respiratory tract Gram-positive bacterial infection (PR=1.55).

Table 2. Observed Prevalence Ratios and Prevalence Values for the 7 Selected CF-Related Clinical Conditions

The 7 Selected CF-Related Clinical Conditions	Prevalence Ratio (PR)	Prevalence
Pancreatic Insufficiency	4.56	0.07
Acute/Recurrent/Chronic Pancreatitis	2.49	0.06
Family History of CFTR Gene Variant	2.41	0.10
Respiratory Tract Gram-Negative Bacterial Infection	1.02	0.51
Respiratory Tract Gram-Positive Bacterial Infection	1.55	0.31
NTM (MAC or non-MAC)	0.70	0.29
Bronchiectasis and/or Tree-in-Bud Opacities	0.70	0.67

Legend: List of observed prevalence ratios as well as prevalence for the 7 CF-related clinical conditions and composite conditions used in the logistic regression analysis.

The results that follow were obtained by applying to Dataset 1 the multivariate LR methods described with the rule that the patients referred for CF testing are those for whom the inequality $S_A \geq C_0$ is satisfied.

Table 3 lists, in decreasing order of AUC value, the combinations of 3 CF-related clinical conditions with AUC values > 0.7. Each of the 7 selected CF-related clinical conditions appears in Table 3 at least once.

Application of the screening model to the Dataset 1 cohort showed the highest AUC value for the combination of pancreatic insufficiency, acute/chronic pancreatitis, and bronchiectasis or tree-in-bud opacities (AUC=0.786). This was closely followed by the combination of pancreatic insufficiency, bronchiectasis or tree-in-bud opacities, and respiratory tract Gram-negative bacterial infection (AUC=0.770). Application of the

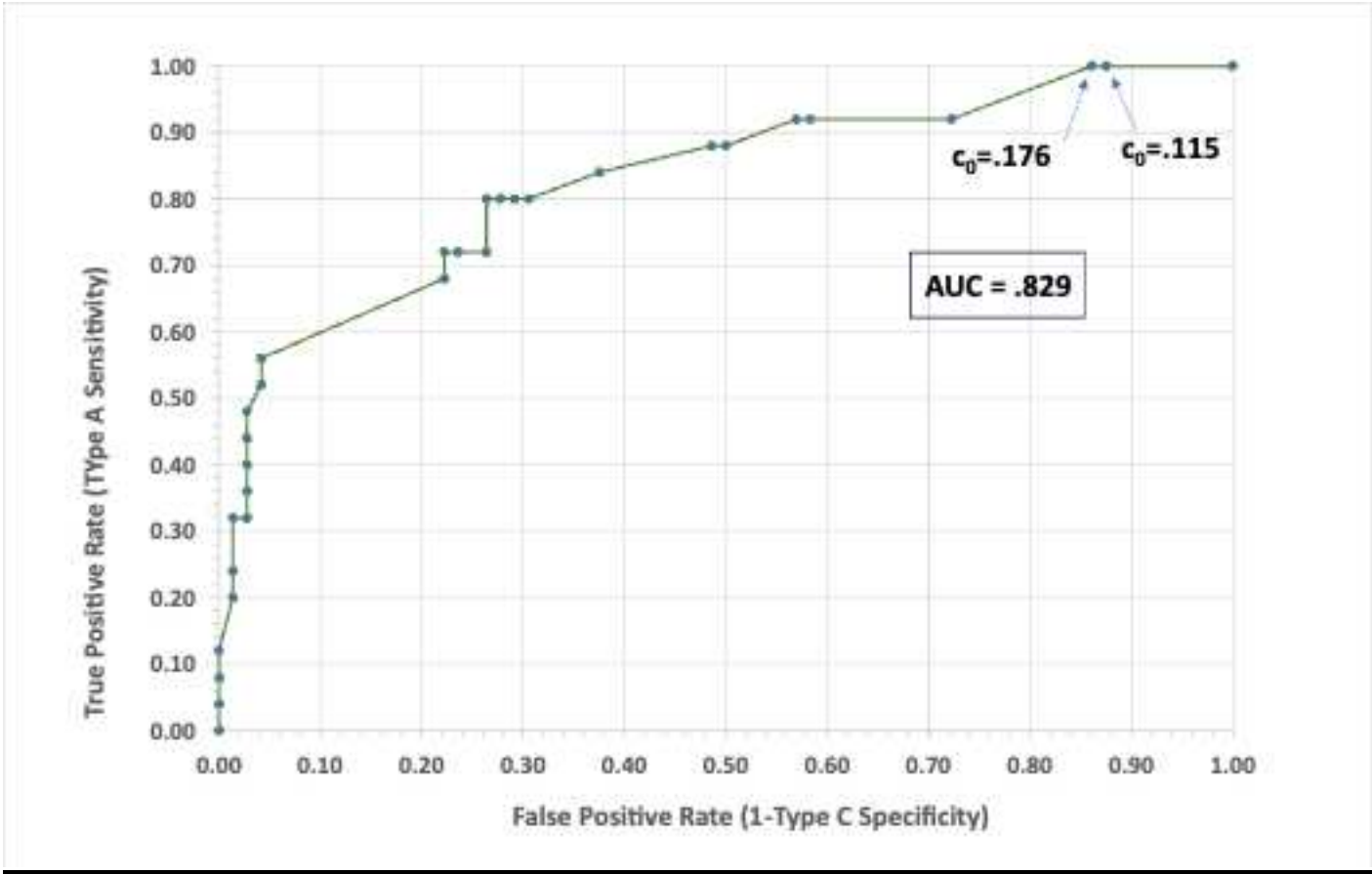
315 screening model to the 171-patient Dataset 1 cohort resulted in the ROC curve shown in Figure 2, with an
 316 AUC value of 0.829.

317 **Table 3. List of 3-variate LR models with AUC Values in Excess of 0.700**

CF-Related Clinical Predictor 1	CF-Related Clinical Predictor 2	CF-Related Clinical Predictor 3	AUC Value
pancreatic insufficiency	a/c pancreatitis	bronchiectasis and/or tree-in-bud	0.786
pancreatic insufficiency	bronchiectasis or tree-in-bud	gram-negative bacteria	0.770
pancreatic insufficiency	a/c pancreatitis	gram-negative bacteria	0.758
pancreatic insufficiency	bronchiectasis or tree-in-bud	gram-positive bacteria	0.754
pancreatic insufficiency	bronchiectasis or tree-in-bud	NTM (MAC or non-MAC)	0.750
pancreatic insufficiency	family history	gram-negative bacteria	0.748
pancreatic insufficiency	family history	gram-positive bacteria	0.743
pancreatic insufficiency	family history	NTM (MAC or non-MAC)	0.734
pancreatic insufficiency	family history	bronchiectasis or tree-in-bud	0.731
family history	bronchiectasis and/or tree-in-bud	gram-negative bacteria	0.727
pancreatic insufficiency	a/c pancreatitis	gram-positive bacteria	0.722
a/c pancreatitis	bronchiectasis or tree-in-bud	gram-negative bacteria	0.716
pancreatic insufficiency	gram-negative bacteria	gram-positive bacteria	0.713
pancreatic insufficiency	gram-negative bacteria	NTM (MAC or non-MAC)	0.712
a/c pancreatitis	family history	gram-negative bacteria	0.708
family history	bronchiectasis and/or tree-in-bud	gram-positive bacteria	0.701

318
 319 **Legend:** List of 3 variate LR models with AUC in excess of .70.

330 **Figure 2.** ROC Curve Resulting from Application of the Screening Model to Dataset 1



331

332 **Legend:** Application of the screening model to the 171-patient Dataset 1 cohort resulted in the ROC curve
333 shown above with an AUC value of 0.829.

334

335 Both databases include only those Type C patients recommended for CF testing based on clinical judgment.
336 Without the inclusion in the database of the Type C patients not recommended for CF testing based on clinical
337 judgment, the calculated value for the screening model Type C specificity is artificially low. Consequently,
338 the Bootstrap method was not applied to quantify the error in the estimated Type C specificity value.

339

340 Tables 4 and 5 show the sensitivity and specificity values obtained by applying the screening model with the
341 decision rule that a patient is recommended for CF testing based on the LR analysis, i.e. a patient is

recommended for CF testing only if Criterion 1 is satisfied. Table 4 applies to Dataset 1 and Table 5 applies to Dataset 2.

For each dataset, the number (n) of patients is indicated for Type *A*, Type *B*₀, Type *B*₁, and Type *C*. The *C*₀ value of 0.115 is ruled out because the screening model would recommend all 171 Dataset 1 patients for CF testing. As a result, the choice *C*₀=0.176 is made.

If LR analysis and Criterion (1) alone are applied by the screening model, the Type *A* sensitivity value is 1.00 for both Dataset 1 and Dataset 2. A Type *B* patient that tests positive for one CF gene mutation is classified as Type *B*₁ and classified as Type *B*₀ otherwise. For both datasets and both Type *B*₁ and Type *B*₀ patients, the sensitivity value is less than 1.00. Bootstrapping was then used to generate sampling distributions from which the mean values and 95% CIs were obtained.

Table 4. LR Analysis Sensitivity and Specificity Results

Dataset 1

<i>C</i> ₀ Value	Type <i>A</i> (n=25)	Type <i>B</i> ₁ (n=28)	Type <i>B</i> ₀ (n=46)	Type <i>C</i> (n=72)	AUC
	Sensitivity	Sensitivity Mean [95% CI]	Sensitivity Mean [95% CI]	Specificity Mean [95% CI]	Mean [95%CI]
0.176	1.000	0.969 [.903,.975]	0.820 [.675,1.00]	0.139 [.058,.219]	0.826 [.703,.929]

Legend: LR Analysis Sensitivity, Specificity and AUC Results for Dataset 1

Table 5. LR Analysis Sensitivity and Specificity Results

Dataset 2

<i>C₀</i> Value	Type <i>A</i> (n=4)	Type <i>B₁</i> (n=11)	Type <i>B₀</i> (n=16)	Type <i>C</i> (n=54)	AUC
	Sensitivity	Sensitivity Mean [95% CI]	Sensitivity Mean [95% CI]	Specificity Mean [95% CI]	Mean [95%CI]
0.176	1.000	0.962 [.885,1.00]	0.848 [.730,.951]	0.156 [.040,.391]	0.835 [.735,.913]

Legend: LR Analysis Sensitivity, Specificity and AUC Results for Dataset 2

The tabulated results indicate the following: (1) for both Dataset 1 and Dataset 2, the Type *A* sensitivity value is 1.000; (2) for both Dataset 1 and Dataset 2, the sensitivity value for the Type *B₁* patients is higher than the sensitivity value for the Type *B₀* patients; (3) the specificity value for the Dataset 2 Type *C* patients is higher than the specificity value for the Dataset 1 Type *C* patients.

If the patients recommended for CF testing are those that satisfy either Criterion (1) or Criterion (2), Bootstrapping was not required to generate sampling distributions for Type *A* sensitivity, Type *B₀* sensitivity, or Type *B₁* sensitivity, since in all three cases the sensitivity value =1.00 for both Dataset 1 and Dataset 2. In the case of Type *C*, the number of true negatives for each dataset was too small to make it useful from a statistical standpoint to employ Bootstrapping to each dataset to generate a sampling distribution and to determine a 95% confidence interval for Type *C* specificity. The Bootstrapping results for Dataset 1, including the sampling distribution and the 95% confidence intervals for the sensitivity, specificity, and AUC values are tabulated on the Bootstrapping Stats sheet of the Development Evaluation Excel workbook (Additional File 3). The Bootstrapping results for Dataset 2, including the sampling distribution and the 95% confidence intervals for the sensitivity, specificity, and AUC values are tabulated on the Bootstrapping Stats sheet of the Temporal Evaluation Excel workbook (Additional File 4).

383

384 **4. Discussion**

385 The study results indicate that for both Dataset 1 and Dataset 2, the screening model matched the performance
386 of clinical judgment in recommending for CF testing both all Type *A* patients and all Type *B* patients
387 recommended for CF testing based on clinical judgment. In addition, the screening model converted to true
388 negatives 1.59% (2/126) of the Type *C* patients who were false positives based on clinical judgment.

389 The study results support the development of a probabilistic screening model based on a set of selected patient
390 clinical characteristics to assist clinicians in determining which patients to test for CF. The developmental test
391 results indicate that the screening model described performed well in recommending for CF testing those
392 patients that testing classifies as CF as well as patients that CF testing classifies as CFNR. The temporal
393 evaluation results provide further evidence that the screening model can recommend for CF testing patients
394 classified as CF or CFNR who warrant CF testing. .

395 The predictive accuracy results are significantly different when the screening model is applied to Dataset 1
396 and Dataset 2 with the decision rule that a patient is recommended for CF testing if either Criterion 1 or
397 Criterion 2 is satisfied. It is then the case that for both Dataset 1 and Dataset 2, all Type *B* patients as well as
398 all Type *A* patients are recommended for CF testing. Further, for both Dataset 1 and Dataset 2, with few
399 exceptions, the Type *C* patients that fail to satisfy Criterion 1 do satisfy Criterion 2. Consequently, for both
400 datasets, the Type *C* specificity value based on Criterion 1 and Criterion 2 in combination is quite low.

401 Dataset 2 contained only four Type *A* patients; therefore, the finding that all four were recommended for CF
402 testing should be interpreted cautiously because of the small number of Type *A* outcomes. Importantly,
403 however, the final screening rule also recommended CF testing for all 27 Type *B* patients in Dataset 2 (12
404 Type *B*1 and 15 Type *B*0), providing additional evidence that the model identified patients for whom CF
405 remained established or unresolved rather than relying solely on the four Type *A* cases.

406 Both databases include only those Type *C* patients recommended for CF testing based on clinical judgment.
407 Without the inclusion in the database of the Type *C* patients not recommended for CF testing based on clinical

408 judgment, the calculated value for the screening model Type C specificity is artificially low. Consequently,
409 the Bootstrap method was not applied to quantify the error in the estimated Type C specificity value.

410

411 Both Datasets 1 and 2 indicate that patients satisfying Criterion 2 can be either Type B or Type C. The study
412 results indicate that achieving a sensitivity value of 1.00 in recommending Type B patients for CF testing
413 requires the recommendation for CF testing of those Type C patients who satisfy Criterion 2.

414

415 The study results support the conclusion that with further testing, the screening model described can be
416 developed into a computer application with a user-friendly interface that is immediately accessible to clinicians
417 with varying levels of experience in the diagnosis of CF. This application would supplement clinical judgment
418 in determining when sweat chloride and CFTR gene variant testing for CF are appropriate. The authors believe
419 this to be the first study to demonstrate a systematic approach in using clinical features to create such a
420 screening model. Prior work has suggested the use of combinations of sweat chloride, genetic screening, and
421 nasal potential difference (NPD) in the diagnosis of CF (21) among adults with bronchiectasis. However, sweat
422 chloride and NPD testing are not universally available. Therefore, the use of a practical model incorporating
423 commonly obtainable diagnostic elements to determine which patients to send for these tests will be of
424 significant clinical utility. Furthermore, because a significant number of adult patients ultimately diagnosed
425 with CF or CFTR-RD have atypical variants and borderline sweat chloride levels, a computer or artificial
426 intelligence application that uses the entire clinical picture will prove valuable in the diagnosis.

427 As indicated in the diagnostic flow diagram in Figure 1, the first step in diagnosis of CF is to obtain a sweat
428 test. As a minimally invasive and relatively inexpensive test, the threshold to obtain sweat tests in adults with
429 the outlined unexplained symptoms should be very low. However, if sweat test results are intermediate (30–
430 59 mmol/L) or diagnostic of CF (≥ 60 mmol/L), or if CF is still felt to be a possibility despite normal sweat test
431 results, with no clear alternative diagnoses, it is appropriate to pursue complete CFTR gene sequencing with
432 the inclusion of intronic deletions. Limited CFTR variant panels are going to be of less value in this population

with higher probability of rare variants. Making sweat testing more accessible in all geographic settings (and potentially remote testing in the future) may have a significant impact on provider/patient willingness to complete testing. The cost of missing a CF diagnosis is not acceptable, and late diagnoses of CF are difficult to make, even for CF experts. Referral to a CF program should be best practice with diagnostic uncertainty and hopefully can be accomplished more easily with the availability of telehealth. The screening model is intended for use by clinicians evaluating adults in whom CF or CFTR-related disease may be considered. Users should have sufficient clinical expertise to accurately identify the specified predictor variables and interpret the model output as an aid to, rather than a replacement for, clinical judgment.

This study has several strengths. Importantly, it introduces and takes the first steps of early temporal evaluation of a screening model that uses clinical variables to determine the likelihood of CF. The only other model that currently exists for comparison to our knowledge is a predictive clinical score in a French population that may help identify adult patients with bronchiectasis who are at higher risk of having CF or PCD (22), but it is not CF specific. The robustness of the model was not affected if patients without CFTR variant testing were excluded. The screening model is based on actual CF clinic data that are available to most clinicians, irrespective of their resources and comfort with diagnosing CF. The screening model has been developed with a deliberate effort to focus on predictive clinical factors. An illustrative example of application of the screening model to a patient is provided in Supplement Section S.6.

5. Study Limitations

This study has multiple notable limitations, including the small sample size, the single center nature of the study and referral from within CF program. Thus our study lacks generalizability to other clinical or geographic settings (geographic bias) and creates referral bias. There is also lack of ethnic and racial diversity which precludes meaningful assessment of model performance across demographic subgroups. Therefore, model fairness cannot be established from this study and should be evaluated in larger, more diverse populations before broad clinical implementation. There is also temporal bias related to the different time

458 periods between the development and temporal evaluation cohorts. See the Supplement for further
459 explanations of limitations and biases in the study.

460 **6. Conclusions**

461 This study provides preliminary evidence that routinely available clinical variables may be used through a
462 prediction model to identify individuals with late-diagnosed cystic fibrosis. Our study achieved the objective
463 of designing, implementing, and testing a screening model to determine whether or not to recommend CF
464 testing for adult patients, with the intent that the screening model could ultimately be used by both clinicians
465 familiar and those not familiar with diagnosing CF. In summary, an algorithm was developed that is based on
466 predictive clinical factors for the diagnosis of CF and CFNR that is feasible and reproducible and should be
467 further developed in a prospective multicenter study with a more diverse patient population to evaluate its
468 potential utility in determining which patients should be tested for CF.

469

470 **Abbreviation List**

471	ABPA	Allergic bronchopulmonary aspergillosis
472	A/c	Acute, recurrent and/or chronic
473	AUC	Area under receiver operating characteristic curve
474	BCC	<i>Burkholderia cepacia</i> complex
475	BMI	Body mass index
476	CF	Cystic Fibrosis
477	CFNR	cystic fibrosis not resolved
478	CFTR	Cystic fibrosis transmembrane regulator
479	CFTR-RD	cystic fibrosis transmembrane regulator-related disorder
480	CI	Confidence intervals
481	CT	Computerized Tomography
482	C ₀	ROC curve cutoff probability used in determining whether to recommend CF testing.

483	D	Dataset 1, Development dataset
484	Expanded	Panel of CFTR variant analysis of at least 200 CF variants
485	FEV1	Forced expiratory volume in one second
486	IQR	Interquartile range
487	LR	Logistic regression
488	M	<i>Mycobacterium</i>
489	MAC	<i>Mycobacterium avium</i> complex
490	ML	Maximum likelihood
491	MAR	Missing data at random
492	MNAR	Missing data not at random
493	MRSA	Methicillin-resistant <i>Staphylococcus aureus</i>
494	MSSA	Methicillin-sensitive <i>Staphylococcus aureus</i>
495	Not tested	No CFTR variant analysis
496	Non-Mac	All other mycobacterial species other than <i>mycobacterium avium complex</i>
497	NTM	Nontuberculous <i>mycobacterium</i>
498	PCD	Primary ciliary dyskinesia
499	pwCF	People with cystic fibrosis
500	PPV	Positive predictive value
501	ROC	Receiver operating characteristic
502	SP	Sample proportion
503	Specific	10 or less CFTR variants tested
504	Spp.	Species
505	Standard	Panel of CFTR variant analysis between 23-106 variants
506	TP	True positive
507	V	Dataset 2, Temporal evaluation dataset

508 VVCC Variant of varying clinical significance

509

510 **Declarations**

511 • **Ethics approval and consent to participate**

512 This retrospective study was approved by the Santa Barbara Cottage Hospital Institutional Review Board (IRB
513 #16-94; approved February 15, 2017). The requirement for informed consent and HIPAA authorization was
514 waived by the Institutional Review Board because of the retrospective nature of the study. The study was
515 conducted in accordance with applicable institutional and ethical guidelines.

516 • **Consent for publication**

517 Not applicable.

518 • **Availability of data and materials**

519 All data generated or analyzed during this study are included in this published article and its supplementary
520 information files. The deidentified development and temporal evaluation datasets are provided in Additional
521 file 2. The model calculations, formulas, logistic regression analyses, ROC analyses, and bootstrap results are
522 provided in Additional files 3 and 4.

523 • **Competing interests**

524 The authors declare that they have no competing interests.

525 • **Funding**

526 This research received no specific grant from any funding agency in the public, commercial, or not-for-profit
527 sectors.

528 • **Authors' contributions**

529 R.B. and Y.M. conceived the study and contributed to the study design and clinical data acquisition. R.B.
530 determined outcome classification. R.B. and Y.M. contributed to interpretation of the results, and preparation
531 and revision of the manuscript. B.B. contributed to study design, development and implementation of the

532 statistical methodology and logistic regression model, statistical analysis, interpretation of the results, and
533 preparation and revision of the manuscript. R.B. is the guarantor of the manuscript; he always had full access
534 to the data from the study and takes responsibility for the integrity of all data and analyses.

535 R.B., Y.M., and B.B. wrote the first draft of the manuscript. R.B. and B.B. wrote the final draft of the
536 manuscript. All authors approved the final version of the manuscript. The preliminary results of this study
537 were presented in abstract form at the North American Cystic Fibrosis Conference in November 2021 and
538 September 2024. There are no financial or non-financial disclosures. There were no sponsors of this study.

539

540 • **Acknowledgements**

541 The authors thank the clinical and research staff who contributed to data collection and maintenance of the
542 clinical databases used in this study including Joe Carmona, BS, Cori Stritzel, RN, Dominique Rosuello, Mony
543 Son, MD, Ginerosa Carbone, MD, Austin Moser, MD. We also thank Amelia Jones, MS, CCRP for assistance
544 with preparation of the manuscript.

545

546 Additional files

547 **Additional file 1**

548 File name: Additional_file_1_Supplementary_Material.docx; File format: DOCX

549 Title: Supplementary Methods and Data Tables; Description: Detailed description of the multi-model logistic
550 regression analysis, model application, missing-data handling, study limitations, and supplementary tables
551 for the development and temporal evaluation cohorts.

552 **Additional file 2**

553 File name: Additional_file_2_Deidentified_Datasets.xlsx; File format: XLSX

554 Title: Deidentified Development and Temporal Evaluation Datasets; Description: Deidentified patient-level
555 data for Dataset 1, the development cohort, and Dataset 2, the temporal evaluation cohort.

556 **Additional file 3**

557 File name: Additional_file_3_LR_Development_Workbook.xlsx; File format: XLSX

558 Title: Logistic Regression Development Workbook; Description: Self-contained Excel workbook containing
559 the model-development calculations, logistic regression models, ROC analysis, model-performance
560 calculations, and bootstrap results for Dataset 1.

561 **Additional file 4**

562 File name: Additional_file_4_LR_Temporal_Evaluation_Workbook.xlsx; File format: XLSX

563 Title: Logistic Regression Temporal Evaluation Workbook; Description: Self-contained Excel workbook
564 containing the temporal evaluation calculations, model-performance analysis, and bootstrap results for
565 Dataset 2.

References

- 566 1. Ziedalski TM, Kao PN, Henig NR, Jacobs SS, Ruoss SJ. Prospective Analysis of Cystic
567 Fibrosis Transmembrane Regulator Mutations in Adults With Bronchiectasis or
568 Pulmonary Nontuberculous Mycobacterial Infection. Chest. 2006 Oct;130(4):995–1002.
569 doi:10.1378/chest.130.4.995

2. Chalmers JD, Haworth CS, Flume P, Long MB, Burgel PR, Dimakou K, et al. European Respiratory Society clinical practice guideline for the management of adult bronchiectasis. *European Respiratory Journal*. 2025 Dec;66(6):2501126. doi:10.1183/13993003.01126-2025
3. T Hill A, L Sullivan A, D Chalmers J, De Soyza A, Stuart Elborn J, Andres Floto R, et al. British Thoracic Society Guideline for bronchiectasis in adults. *Thorax*. 2019 Jan;74(Suppl 1):1–69. doi:10.1136/thoraxjnl-2018-212463
4. Nick JA, Nichols DP. Diagnosis of Adult Patients with Cystic Fibrosis. *Clinics in Chest Medicine*. 2016. doi:10.1016/j.ccm.2015.11.006
5. Keating CL, Liu X, DiMango EA. Classic respiratory disease but atypical diagnostic testing distinguishes adult presentation of cystic fibrosis. *Chest*. 2010;137(5):1157–63. doi:10.1378/chest.09-1352
6. Gilljam M, Ellis L, Corey M, Zielenski J, Durie P, Tullis DE. Clinical Manifestations of Cystic Fibrosis Among Patients With Diagnosis in Adulthood. *Chest*. 2004 Oct;126(4):1215–24. doi:10.1378/chest.126.4.1215
7. Castellani C, Linnane B, Pranke I, Cresta F, Sernmet-Gaudelus I, Peckham D. Cystic Fibrosis Diagnosis in Newborns, Children, and Adults. *Semin Respir Crit Care Med*. 2019 Dec 3;40(06):701–14. doi:10.1055/s-0039-1697961
8. JUNG FW, MACRAE S. NONCYSTIC FIBROSIS BRONCHIECTASIS IN US PRIMARY CARE PATIENTS: PREVALENCE AND INCIDENCE FROM 2013-2024. *Chest*. 2025 Oct;168(4):A194. doi:10.1016/j.chest.2025.07.114
9. Farley H, Poole S, Chapman S, Flight W. Diagnosis of cystic fibrosis in adulthood and eligibility for novel CFTR modulator therapy. *Postgrad Med J*. 2022 May 1;98(1159):341–5. doi:10.1136/postgradmedj-2020-139278

10. Middleton PG, Mall MA, Dřevínek P, Lands LC, McKone EF, Polineni D, et al.
Elexacaftor–Tezacaftor–Ivacaftor for Cystic Fibrosis with a Single Phe508del Allele.
New England Journal of Medicine. 2019 Nov 7;381(19):1809–19.
doi:10.1056/NEJMoa1908639
11. Grasemann H, Ratjen F. Cystic Fibrosis. New England Journal of Medicine. 2023 Nov
2;389(18):1693–707. doi:10.1056/NEJMra2216474
12. Burgel PR, Burnet E, Regard L, Martin C. The Changing Epidemiology of Cystic
Fibrosis. Chest. 2023 Jan;163(1):89–99. doi:10.1016/j.chest.2022.07.004
13. Langfelder-Schwind E, Karczeski B, Strecker MN, Redman J, Sugarman EA, Zaleski C,
et al. Molecular Testing for Cystic Fibrosis Carrier Status Practice Guidelines:
Recommendations of the National Society of Genetic Counselors. J Genet Couns. 2014
Feb 7;23(1):5–15. doi:10.1007/s10897-013-9636-9
14. Widerman E. The Experience of Receiving a Diagnosis of Cystic Fibrosis After Age 20.
Soc Work Health Care. 2005 Feb 8;39(3–4):415–33. doi:10.1300/J010v39n03_12
15. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al.
TRIPOD+AI statement: updated guidance for reporting clinical prediction models that
use regression or machine learning methods. BMJ. 2024 Apr 16;385:e078378.
doi:10.1136/bmj-2023-078378
16. Sosnay PR, White TB, Farrell PM, Ren CL, Derichs N, Howenstine MS, et al.
Diagnosis of Cystic Fibrosis in Nonscreened Populations. J Pediatr. 2017 Feb;181:S52-
S57.e2. doi:10.1016/j.jpeds.2016.09.068
17. 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 Feb;181:S4-S15.e1. doi:10.1016/j.jpeds.2016.09.064

18. Sosnay PR, Salinas DB, White TB, Ren CL, Farrell PM, Raraigh KS, et al. Applying Cystic Fibrosis Transmembrane Conductance Regulator Genetics and CFTR2 Data to Facilitate Diagnoses. *J Pediatr*. 2017 Feb;181:S27-S32.e1. doi:10.1016/j.jpeds.2016.09.063
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009 Apr;42(2):377–81. doi:10.1016/j.jbi.2008.08.010
20. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O’Neal L, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform*. 2019 Jul;95:103208. doi:10.1016/j.jbi.2019.103208
21. Hubert D, Fajac I, Bienvenu T, Desmazes-Dufeu N, Ellaffi M, Dall’Ava-Santucci J, et al. Diagnosis of cystic fibrosis in adults with diffuse bronchiectasis. *Journal of Cystic Fibrosis*. 2004 Mar;3(1):15–22. doi:10.1016/j.jcf.2003.12.004
22. Schlemmer F, Hamzaoui A, Zebachi S, Le Thuaut A, Mangiapan G, Monnet I, et al. Etiological Work-Up for Adults with Bronchiectasis: A Predictive Diagnostic Score for Primary Ciliary Dyskinesia and Cystic Fibrosis. *J Clin Med*. 2021 Aug 6;10(16):3478. doi:10.3390/jcm10163478

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1SupplementaryMaterial.docx](#)
- [Additionalfile2DeidentifiedDatasets.xlsx](#)
- [Additionalfile3LRDevelopmentWorkbook.xlsx](#)
- [Additionalfile4LRTemporalEvaluationWorkbook.xlsx](#)