UNB
EST. 1785
UNIVERSITY OF NEW BRUNSWICK

NB-IRDT
New Brunswick Institute for Research, Data and Training

Technical Appendix

NB-IRDT Chronic Obstructive Pulmonary Disease Research Program – Report Two:

Investigation of the Canadian Chronic Disease Surveillance System (CCDSS) and the New Brunswick COPD Health Information Platform (NB-CHIP)

Kyle Rogers, MSc
Chris Folkins, PhD
Ted McDonald, PhD

Project Title

Technical appendix: NB-IRDT Chronic Obstructive Pulmonary Disease research program – Report two: Investigation of the Canadian Chronic Disease Surveillance System (CCDSS) and the New Brunswick COPD Health Information Platform (NB-CHIP)

Principal Investigator

Chris Folkins, Research Scientist, NB-IRDT
Ted McDonald, Director, NB-IRDT

Research Team

Kyle Rogers, Data Analyst, NB-IRDT

Publication Date

April 2023

Project Number

P0040: NB-CHIP validation and COPD / asthma case-identification pilot study

Acknowledgements

This study was supported by the Maritime SPOR Support Unit (MSSU), which receives financial support from the Canadian Institutes of Health Research (CIHR), the Nova Scotia Department of Health and Wellness, the New Brunswick Department of Health, the Nova Scotia Health Research Foundation (NSHRF), and the New Brunswick Health Research Foundation (NBHRF). Funding for this project was also provided by AstraZeneca and NBHRF as part of a separate funding agreement. The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by MSSU and its associated funders, NBHRF or AstraZeneca is intended or should be inferred.

Project Data

Research analytic outputs were produced using platform data accessed through the New Brunswick Institute for Research, Data and Training.

How to Cite This Product

Rogers, K., Folkins, C., & McDonald, T. (2023). Technical appendix: NB-IRDT Chronic Obstructive Pulmonary Disease research program – Report two: Investigation of the Canadian Chronic Disease Surveillance System (CCDSS) and the New Brunswick COPD Health Information Platform (NB-CHIP). Fredericton, NB: New Brunswick Institute for Research, Data and Training.

Table of Contents

List of Tables

List of Figures

Overview

The objective of this Technical Appendix is to describe the methodology used in the accompanying report: *NB-IRDT Chronic Obstructive Pulmonary Disease research program – Report two: Investigation of the Canadian Chronic Disease Surveillance System (CCDSS) and the New Brunswick COPD Health Information Platform (NB-CHIP)*.

This document is meant to complement the report and provide greater detail in describing the methods, data sets, constructs, and data quality assessments used in the preparation for the full report.

The first section describes in greater detail the methods used to generate the report, while the second section describes some data quality exploration that was undertaken simultaneously as the report was being prepared.

Methods

Data

Canadian Chronic Disease Surveillance System (CCDSS) – COPD

The CCDSS COPD database contains individual-level surveillance data for COPD in New Brunswick.[1] The CCDSS uses administrative databases to provide a passive surveillance of chronic diseases.[2] An individual is flagged as a CCDSS COPD case if they are aged 35 or older and have one physician service billing or one acute care separation reflecting a COPD diagnosis.

Citizen Data

The Citizen Data contains basic demographic and location information on all residents of New Brunswick who have been issued a provincial Medicare card.[3]

Discharge Abstract Data (DAD)

The DAD contains all acute care separations and day surgeries from institutions in New Brunswick or associated with an NB Medicare number.[4] The DAD captures administrative, clinical, and demographic information including discharges, deaths, sign-outs, and transfers within a fiscal year (April 1 to March 31).

New Brunswick COPD Health Information Platform (NB-CHIP) – Horizon

NB-CHIP contains pulmonary function testing (PFT) data collected during regular care practice from nine pulmonary function test facilities across both Regional Health Authorities (Vitalité Health Network and Horizon Health Network) in New Brunswick.[5] NB-CHIP captures demographic, clinical, and health information.

Note: PFTs are not exclusively used for testing COPD; a variety of health issues can trigger patient symptoms and complaints that would lead a health professional to order a PFT. As such, the presence of a record in NB-CHIP is not definitively indicative of COPD.

Table 1: Data Set-Specific Data Preparation Activities

Data Set	Action
CCDSS COPD	Only the earliest record for each individual was considered in the analysis.
Citizen Data	Only active status addresses for each individual that pertained to the time period of interest were maintained.
DAD	The Hospital Risk Frailty score (HFRS) examines all hospitalizations in the DAD associated with an individual over a 2-year accrual period to create a frailty risk score for the current year. The HFRS algorithm was used to generate an HFRS score for the most recent year an individual was hospitalized More information on frailty is found in the Constructs section below.
NB-CHIP	Only the earliest positive PFT record was maintained. If there were no positive PFT records, the first PFT record was maintained.

Coincidence and Concordance Group Specifications

Table 2: Coincidence Groups Describing CCDSS COPD Diagnosis and Presence of a PFT

CCDSS/ NB-CHIP Coincidence	Group Definition
CCDSS, no PFT	Includes any individual with the following two criteria: • A record in the CCDSS data set • No record in NB-CHIP

CCDSS, PFT	Includes any individual with the following two criteria: • A record in the CCDSS data set • A record in NB-CHIP
No CCDSS, PFT	Includes any individual with the following two criteria: • No record in the CCDSS data set • A record in NB-CHIP

Table 3: Concordance Groups Describing CCDSS COPD Diagnosis and the Results of PFTs

CCDSS/ NB-CHIP Concordance	Group Definition
CCDSS case, positive PFT (PFT diagnosed)	Includes any individual with each of the following: • A record in the CCDSS data set • A record in NB-CHIP with pre- and post-bronchodilator values • At least 1 NB-CHIP record with positive test result (*Positive result = PFT showing airflow limitation indicative of COPD, i.e., post-bronchodilator FEV1/FVC < 0.7)
CCDSS case, negative PFT (overdiagnosed)	Includes any individual with each of the following: • A record in the CCDSS data set • A record in NB-CHIP • No NB-CHIP records with positive test results
CCDSS case, missing or incomplete PFT (non-PFT diagnosed)	Includes any individual with both of the following: • A record in the CCDSS data set • No record in NB-CHIP *or* a record in NB-CHIP with no post-bronchodilator values
Not a CCDSS case, positive PFT (underdiagnosed)	Includes any individual with each of the following: • No record in the CCDSS data set • A record in NB-CHIP • At least 1 NB-CHIP record with a positive test result

Constructs

Airflow Limitation

Airflow limitation for individuals who meet PFT criteria for COPD was derived from NB-CHIP data and defined using the Global Initiative for Obstructive Lung Disease (GOLD) 2020 standards.[6] In patients with post-bronchodilator FEV1/FVC values less than 0.7 (FEV1: forced expiratory volume in one second; FVC: forced vital capacity), the following classifications were used:

Classification	Predictions
Mild	FEV1 ≥ 80% predicted
Moderate	50% ≤ FEV1 < 80% predicted
Severe	30% ≤ FEV1 < 50% predicted
Very Severe	FEV1 < 30% predicted

Only NB-CHIP records have airflow values, so analyses examining airflow were limited to the NB-CHIP subset.

Area-Specific Income per Single Person Equivalent Quintiles (QAIPPE)

Area-specific income was derived from a household size-adjusted measure of after-tax household income, based on 2006 Census data. The individual data were aggregated to the dissemination area level.[7] This dissemination area average value was used to rank all dissemination areas, then the population was divided into approximate fifths to establish quintiles. QAIPPE (Quintile of Annual Income Per Person Equivalent) values apply to all records included in the analyses.

Body Mass Index (BMI)

BMI was derived from NB-CHIP and defined via the Health Canada BMI health risk nomogram.[8]

Classification	Body Mass Index (BMI)
Underweight	< 18.5
Normal weight	18.5 - 24.9
Overweight	25.0 - 29.9
Obese	30+

Only NB-CHIP records have data on BMI, so analyses examining BMI were limited to the NB-CHIP subset.

COPD Case

A record was defined as a COPD case if an individual appeared in the CCDSS COPD data set between 2007 and 2017.

Frailty

Frailty was derived from the DAD using the Hospital Frailty Risk Score (HFRS).[9] The HFRS is a validated algorithm that utilizes diagnosis and intervention codes from administrative acute care separations data over a 2-year period to assign a frailty risk score to individuals.

The HFRS methodology was applied to DAD records between fiscal years 2007 and 2017 to generate a frailty risk score for every individual who experienced one or more hospitalizations during this period. Records for individuals aged 75 and older were then linked to the NB-CHIP and CCDSS COPD data sets. The sample was restricted to age 75 years and older because the HFRS was designed to be used in older adults.

Frailty values apply to all records included in this subsample analysis. Individuals who were not present in DAD records (and therefore not hospitalized) during the specified period were not assigned a frailty risk score and were therefore excluded from frailty analyses.

Positive Pulmonary Function Test (PFT+)

PFT positive (PFT+) test results were defined according to GOLD criteria for airflow limitation.[6]

A record was defined as PFT+ when a post-bronchodilator value was present and the post-bronchodilator FEV1/FVC value was less than 70 percent. An individual was considered PFT+ if they had any positive result between 2007 and 2017.

Results are discernable as PFT+ only when a post-bronchodilator test is applied, so analyses examining PFT+ records were limited to the subset of the NB-CHIP data with post-bronchodilator values.

Note: Positive PFT results are indicative of COPD, but not definitively so. A COPD diagnosis requires patient history, symptoms, and differential diagnoses to be considered. As such, a PFT+ result should not be considered a COPD diagnosis.

Negative Pulmonary Function Test (PFT-)

An individual was defined as PFT negative (PFT-) when they had at least one record present in NB-CHIP between 2007 and 2017 and no NB-CHIP records with positive test results.

Results are discernable as PFT- only when a post-bronchodilator test is applied, so analyses examining PFT- records were limited to the subset of the NB-CHIP data with post-bronchodilator values.

Post-Bronchodilator Not Applied

Records were labelled 'post-bronchodilator not applied' when individuals received a PFT with only a pre-bronchodilator test being applied.

Data Quality and Clinical Issues Identified in NB-CHIP

Introduction

NB-CHIP is a novel data set generated from regular clinical practices and is still undergoing data quality checks to confirm its validity of use for research purposes. The presence of significant data quality issues can undermine any results generated from a data set and introduce bias, which makes any results suspect. As such, identifying, ameliorating, and addressing data quality issues in new data sets is crucial to ensuring they are useful data products that can provide accurate and actionable evidence.

One overarching issue identified among records within NB-CHIP was the completeness of PFT records. If a record in NB-CHIP does not contain complete PFT information, it cannot be used to fully describe the measured clinical values of individuals with respiratory complaints who have gone to a clinic for testing.

Two potential issues affecting the completeness of measured values in NB-CHIP records during the analyses for this project were identified. One of these was discovered to be a data quality issue, and the other was more likely a clinical issue:

1. Data quality issue: PFT records missing all airflow measurement data.

2. Clinical issue: PFT records missing post-bronchodilator airflow measurement data.

In the data quality issue – PFT records missing all airflow measurement data – the records contain basic demographic and physiological information such as age, sex, weight, and height, but they do not contain any of the measured airflow information. The reason for this issue is unclear at this time, but preliminary work exploring the issue suggests it may be a system-generated error, and work with the data owners could fix this issue for future analyses using the data.

The other issue – PFT records missing post-bronchodilator airflow measurement – likely represents a clinical issue rather than a data quality issue in most cases. Specifically, these are records wherein care providers appear not to have performed a second PFT maneuver after

administration of a bronchodilator to the patient (which is the recommended approach when using spirometry to diagnose or rule out COPD). These records only contain 'pre-bronchodilator' values and are missing 'post-bronchodilator' values.

There are a variety of reasons why care providers may not obtain post-bronchodilator values:

- The suspected diagnosis did not indicate a post-bronchodilator test.
- The suspected diagnosis did indicate a post-bronchodilator test, but it was not performed for a valid reason.
- The suspected diagnosis did indicate a post-bronchodilator test, but it was not performed for an invalid reason.

In addition to clinical issues, it remains possible that data quality issues may also contribute to the absence of post-bronchodilator measurements in patient records. An example would be a scenario in which the suspected diagnosis did indicate a post-bronchodilator test, and the test was performed, but it either was not recorded or was not captured in the database due to documentation or data processing problems.

The analyses below examine the completeness of PFTs by demographic (age, sex), physiological (BMI, frailty), health system (health zone of PFT clinic), and socioeconomic characteristics (area income of patient residence) to explore potential factors affecting PFT completeness.

PFT completeness is further broken down into the two issues described above (PFTs missing all airflow data, and PFTs missing post-bronchodilator data) to see how each relates to the explanatory variables described above, and how they relate to the overarching PFT completeness issue.

Finally, by linking NB-CHIP records to the CCDSS COPD data, we were able to further explore any trends in explanatory variables among PFTs missing all airflow data and PFTs missing post-bronchodilator data in relation to whether the tested individuals had been diagnosed with COPD.

Results

Age

Figure 1: Completeness of PFT by Age Group

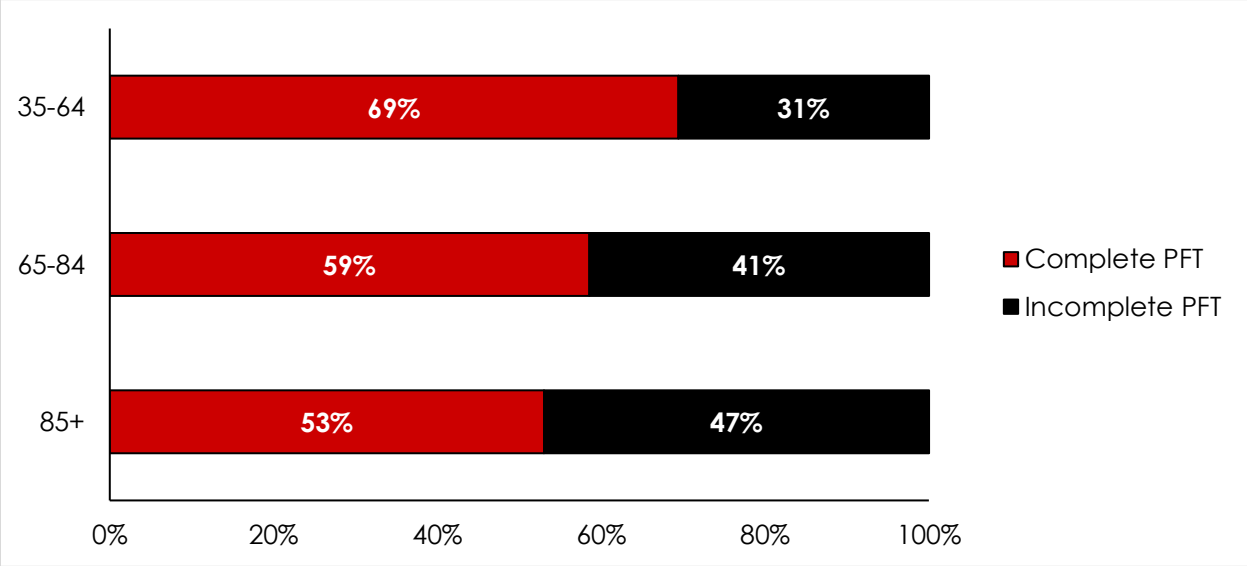


Figure 1 describes the completeness of the pulmonary function tests (PFTs) by age group, with 'Incomplete PFT' capturing both records missing all airflow data due to a data quality issue and any records that do not have a post-bronchodilator value. As seen in Figure 1, the proportion of incomplete PFTs increases with age, going from 31 percent among individuals aged 35 to 64 to 47 percent among individuals aged 85 and over.

Figures 2 to 5 describe the missing airflow data quality issue and lack of post-bronchodilator testing in more detail, presenting their relationship with age among CCDSS cases and non-cases.

Figure 2 describes the missing airflow data issue among CCDSS cases. The proportion of PFT records missing airflow data increases with age, going from 12 percent among individuals aged 35 to 64 up to 20 percent among individuals aged 85 and over.

This trend is not replicated among individuals who had a PFT but were not CCDSS cases (Figure 3): In this group, individuals aged 65 to 84 had the largest proportion of PFT records missing airflow data (26%).

Figure 2: Missing Airflow Data by Age Group, Among CCDSS Cases

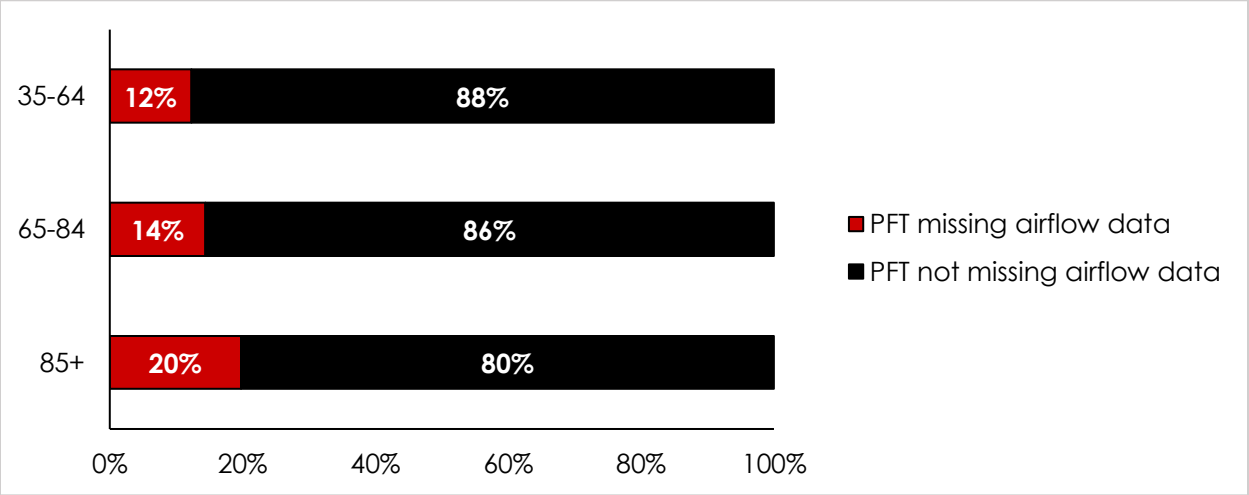


Figure 3: Missing Airflow Data by Age Group, Among Non-CCDSS Cases

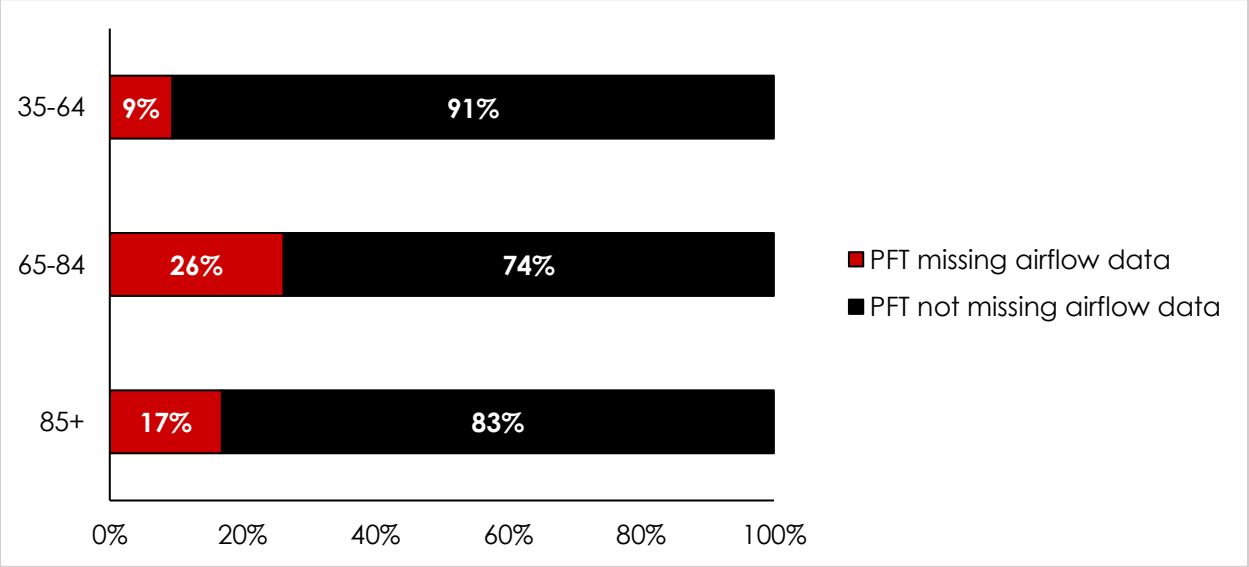


Figure 4 describes the presence of post-bronchodilator values among individuals identified as CCDSS cases. The proportion of PFT records missing a post-bronchodilator test increases with age (13% to 30%).

Figure 5 describes the presence of post-bronchodilator values among individuals who received a PFT but were not CCDSS cases. The proportion of records without a post-bronchodilator value is larger than among CCDSS cases across all age groups. Additionally, while individuals aged 85 and over have a greater proportion of records missing post-bronchodilator values (38%), there appears to be little difference among the other age groups (27% for ages 35 to 64 versus 28% for ages 65 to 84).

Figure 4: Missing Post-Bronchodilator Values by Age Group, Among CCDSS Cases

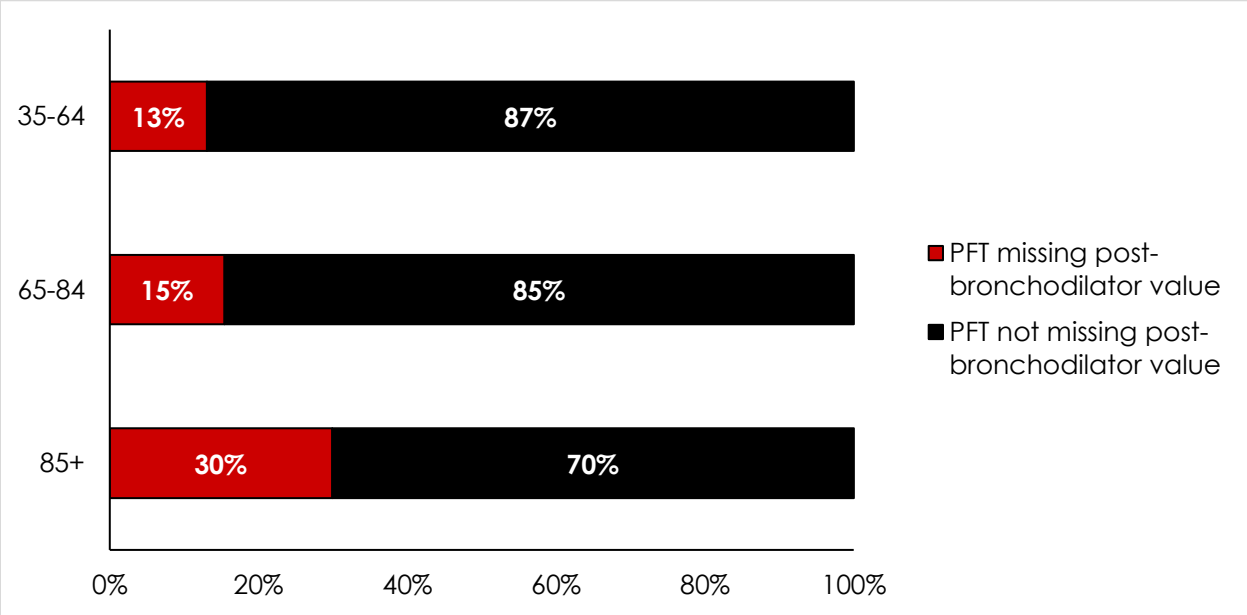
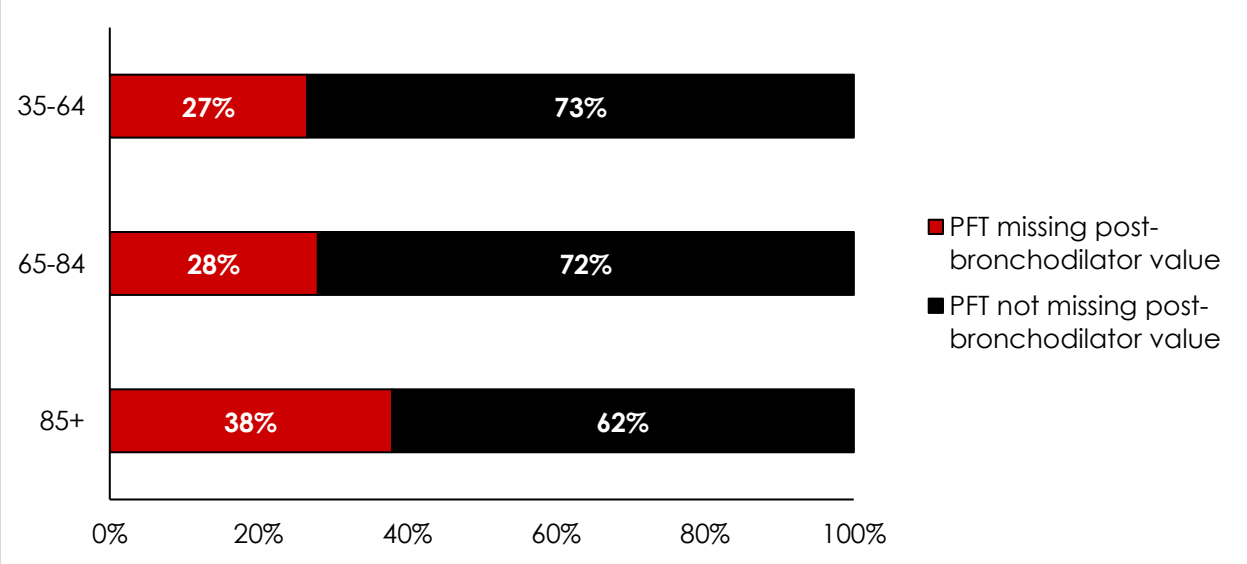


Figure 5: Missing Post-Bronchodilator Values by Age Group, Among Non-CCDSS Cases



Sex

Figure 6 demonstrates that there is very little difference in PFT completeness between males and females, with both having just over one-third of records incomplete owing to either the missing airflow data issue or the lack of post-bronchodilator values.

As seen in Figures 7 to 10, the lack of difference between sexes remains evident when missing airflow data and the absence of post-bronchodilator values are examined individually and among CCDSS cases and non-cases: In all instances, the proportion of incomplete records are similar between males and females.

Figure 6: Completeness of PFT by Sex

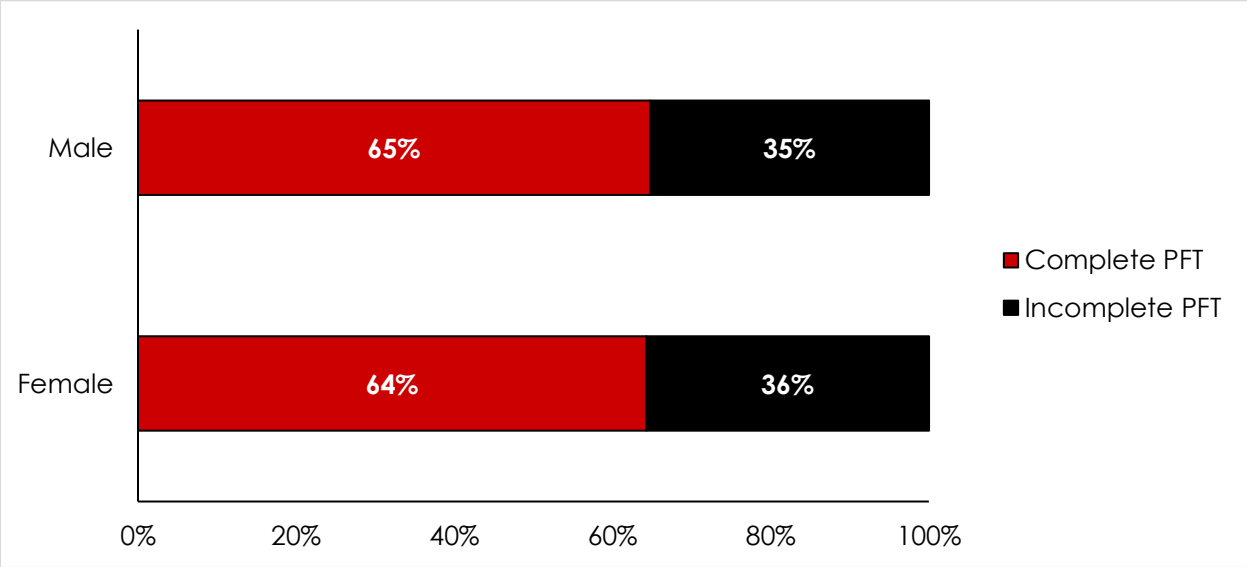


Figure 7: Missing Airflow Data by Sex, Among CCDSS Cases

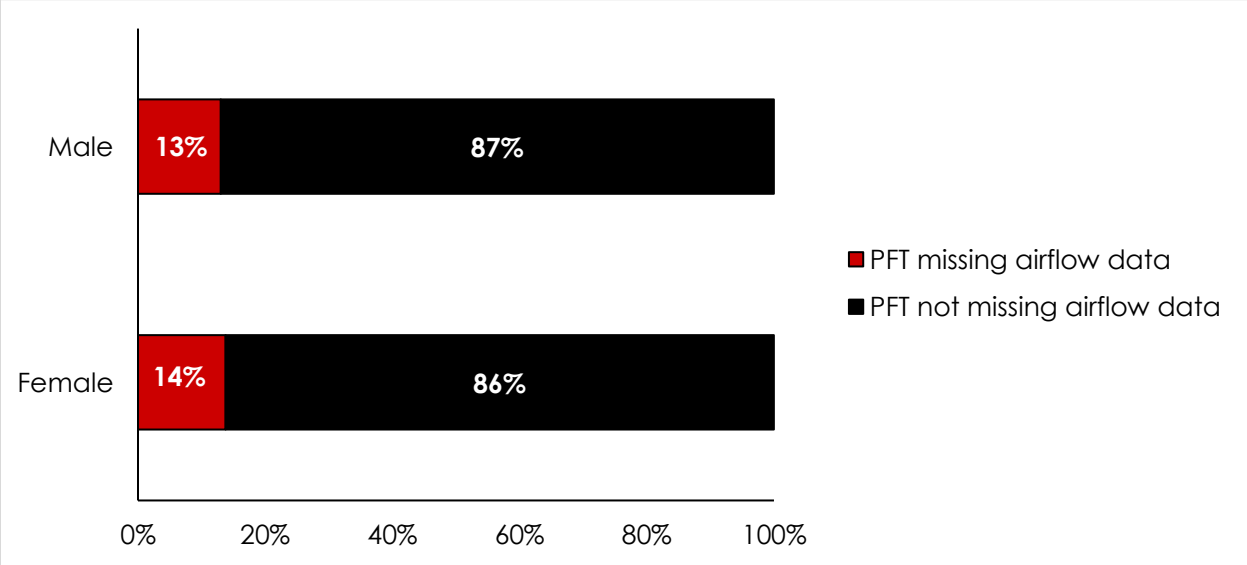


Figure 8: Missing Airflow Data by Sex, Among Non-CCDSS Cases

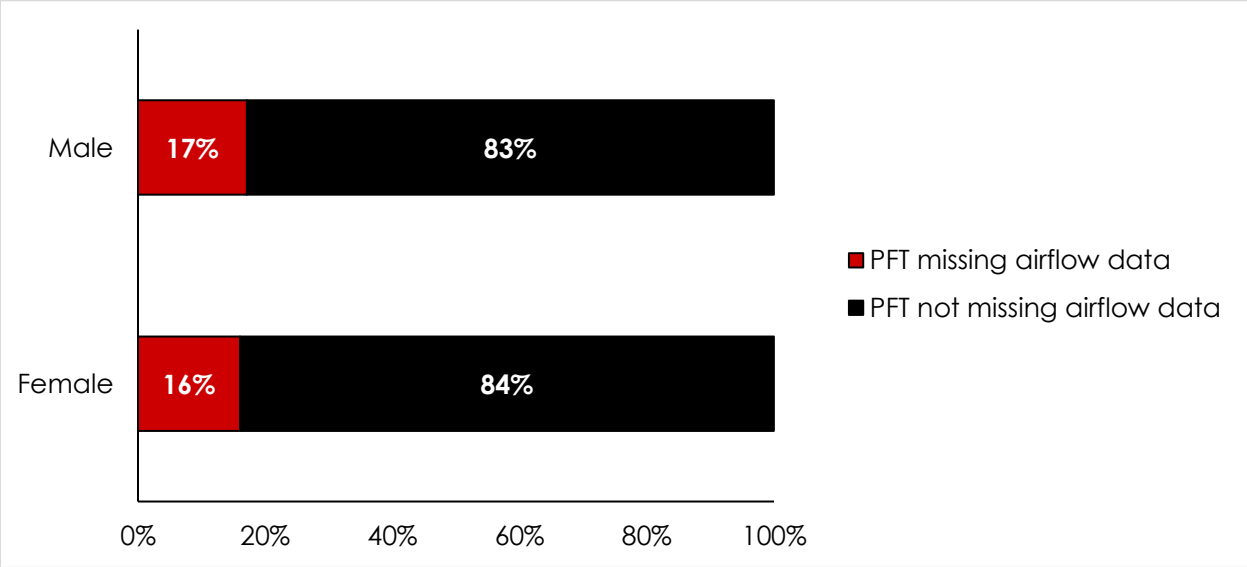


Figure 9: Missing Post-Bronchodilator Values by Sex, Among CCDSS Cases

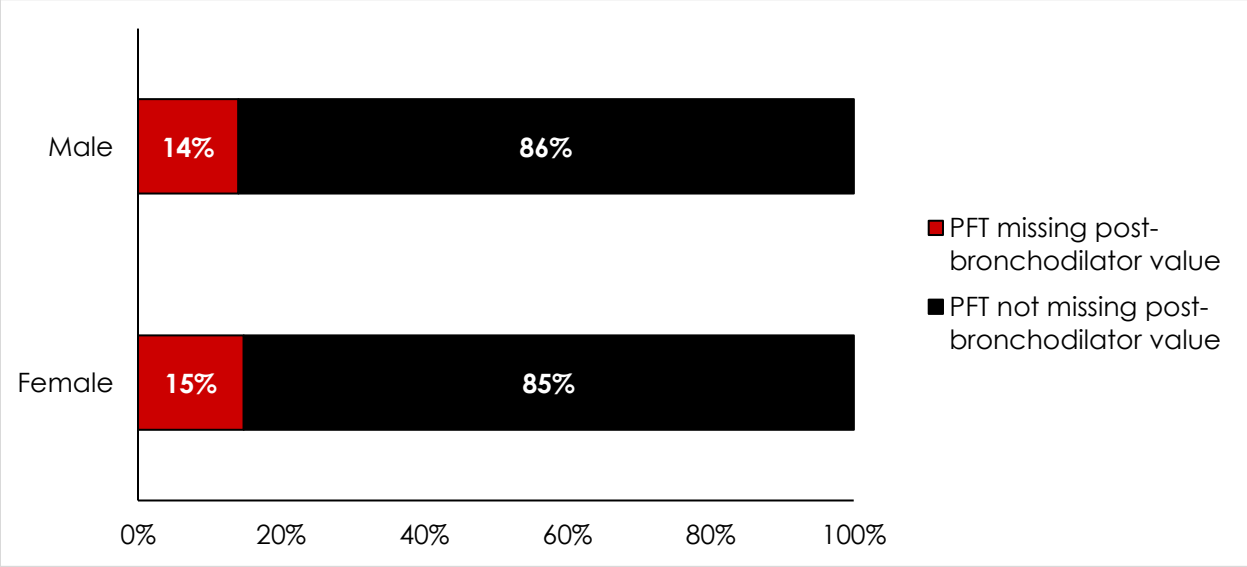
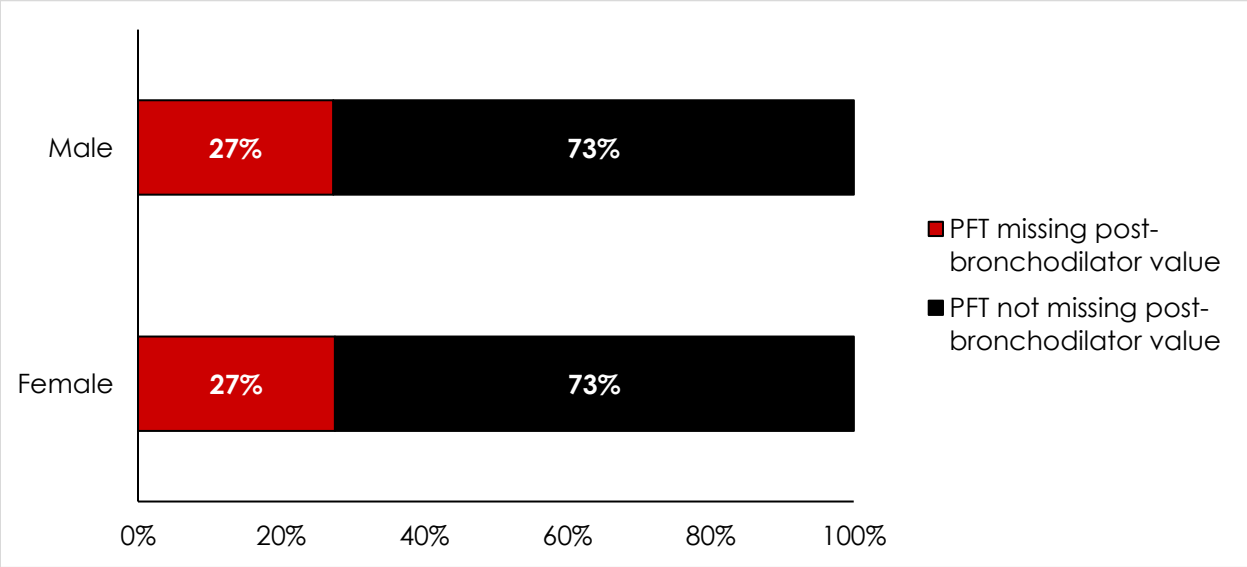


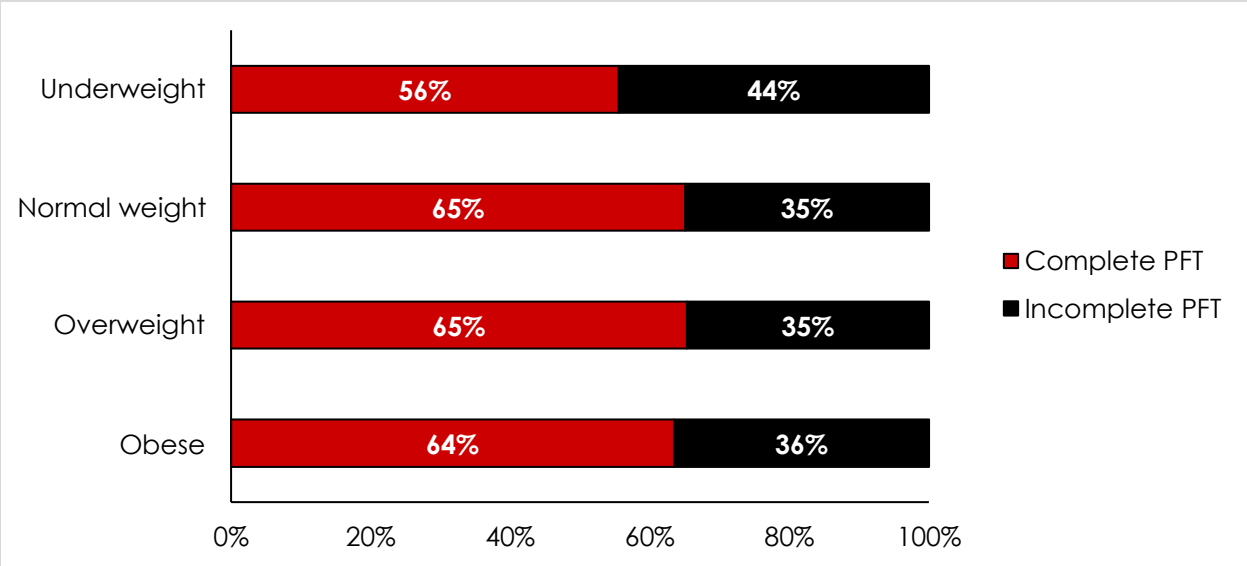
Figure 10: Missing Post-Bronchodilator Values by Sex, Among Non-CCDSS Cases



Body Mass Index (BMI)

Figure 11, which describes PFT completeness in relation to BMI classification, demonstrates little difference in PFT completeness between individuals classified as normal weight, overweight, or obese (64% to 65%). Individuals classified as underweight have a smaller proportion (56%) of complete PFT records compared to other BMI classifications.

Figure 11: Completeness of PFT by BMI Classification



Figures 12 and 13 describe how the missing airflow data quality issue relates to BMI classification, presented by CCDSS status. Among individuals identified as CCDSS cases (Figure 12), 13 to 14 percent of individuals classified as normal weight, overweight, or obese were missing airflow

information from their PFT, while 24 percent of individuals classified as underweight were missing the same information.

Individuals with a PFT but no CCDSS case status (Figure 13) showed a similar trend between BMI categories and a greater proportion of missing airflow data overall across all BMI categories, with 15 to 18 percent of individuals classified as normal weight, overweight, or obese missing airflow data, and more than a third of individuals classified as underweight missing airflow data.

Figure 12: Missing Airflow Data by BMI Classification, Among CCDSS Cases

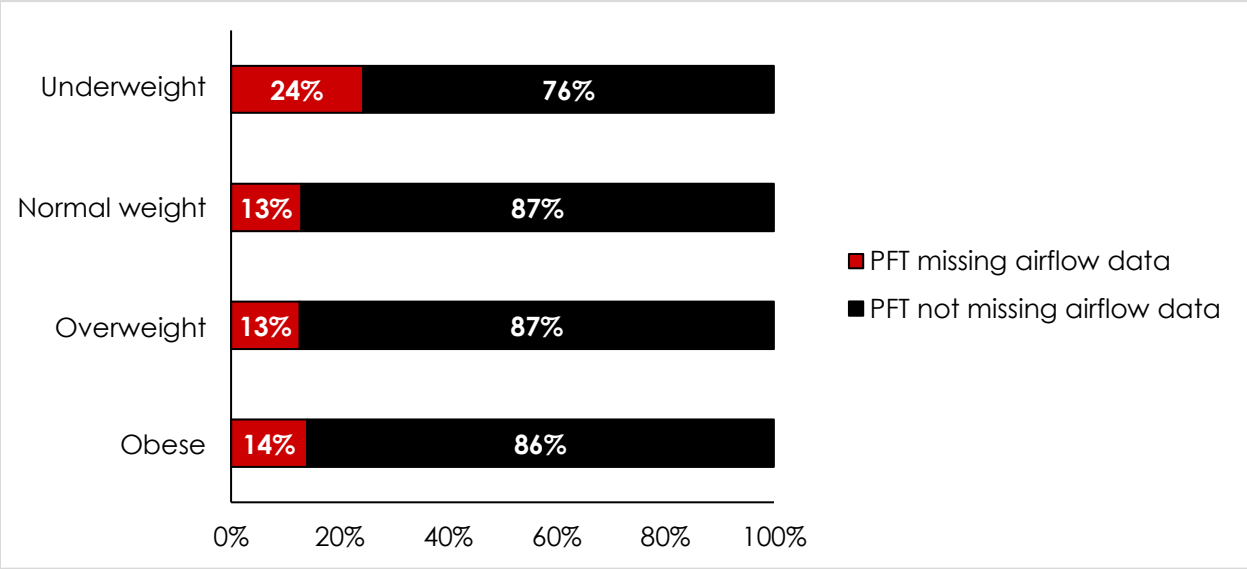
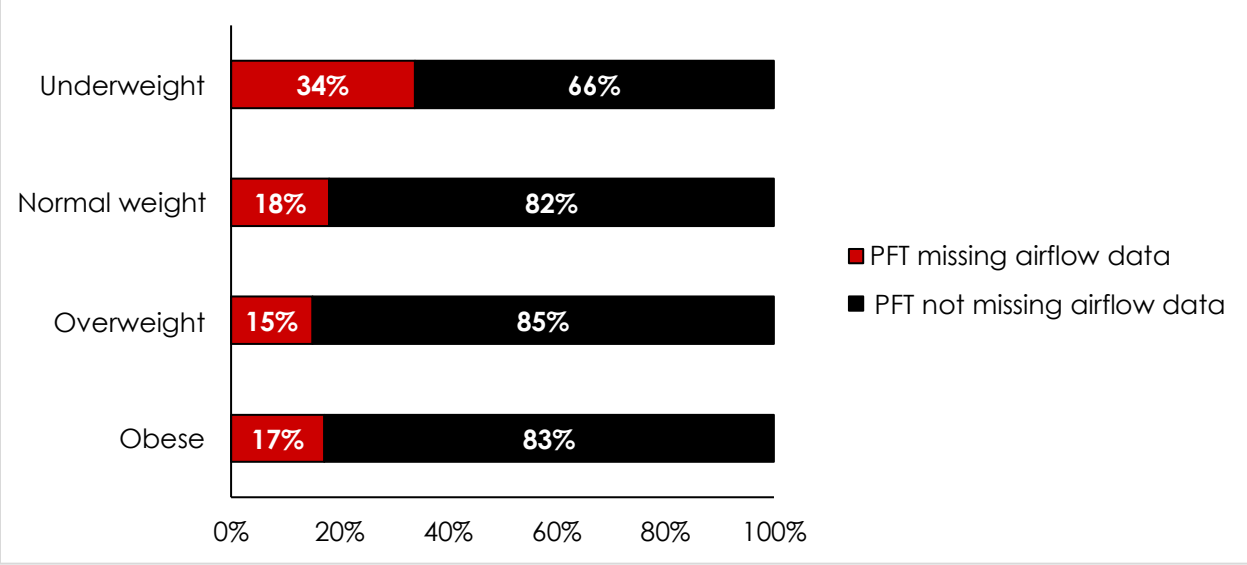


Figure 13: Missing Airflow Data by BMI Classification, Among Non-CCDSS Cases



Figures 14 and 15 describe how the lack of post-bronchodilator values relates to BMI classification, presented by CCDSS status.

Figure 14 demonstrates that there is little variation across BMI classifications among CCDSS cases. A similar trend between BMI categories is observed in Figure 15 among non-CCDSS cases, though with a greater overall proportion of individuals missing post-bronchodilator values across all BMI categories.

Figure 14: Missing Post-Bronchodilator Values by BMI Classification, Among CCDSS Cases

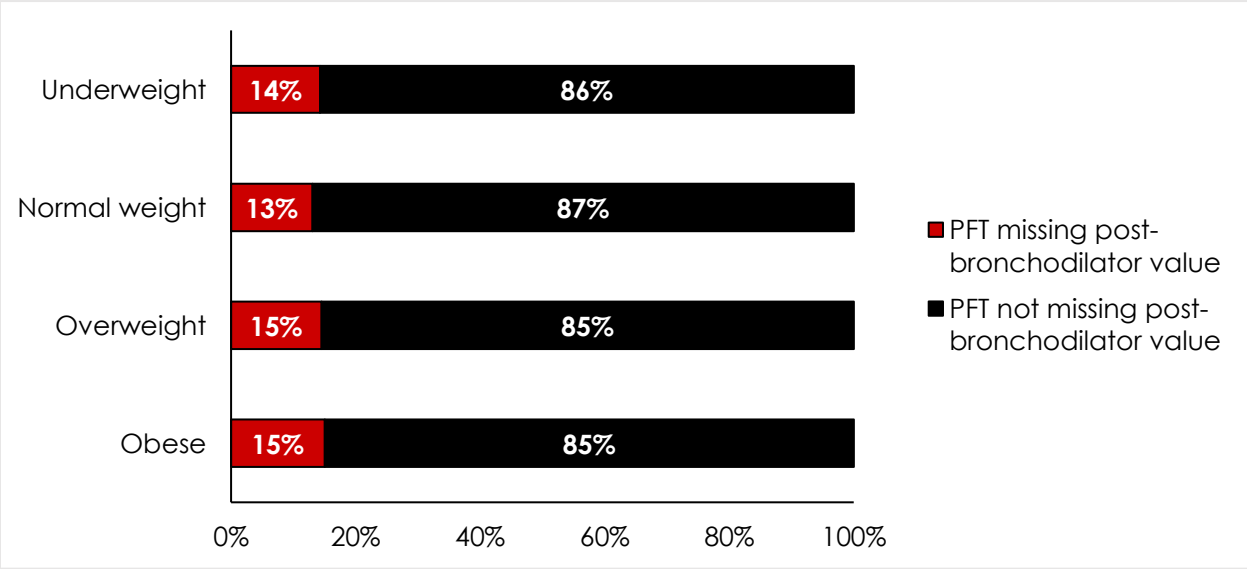
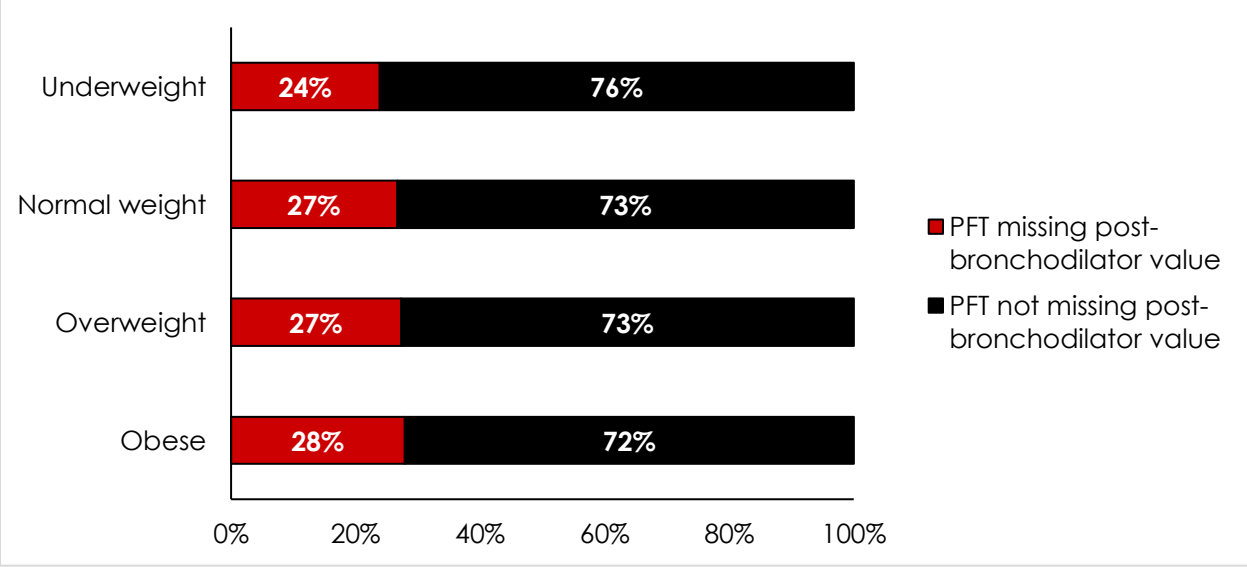


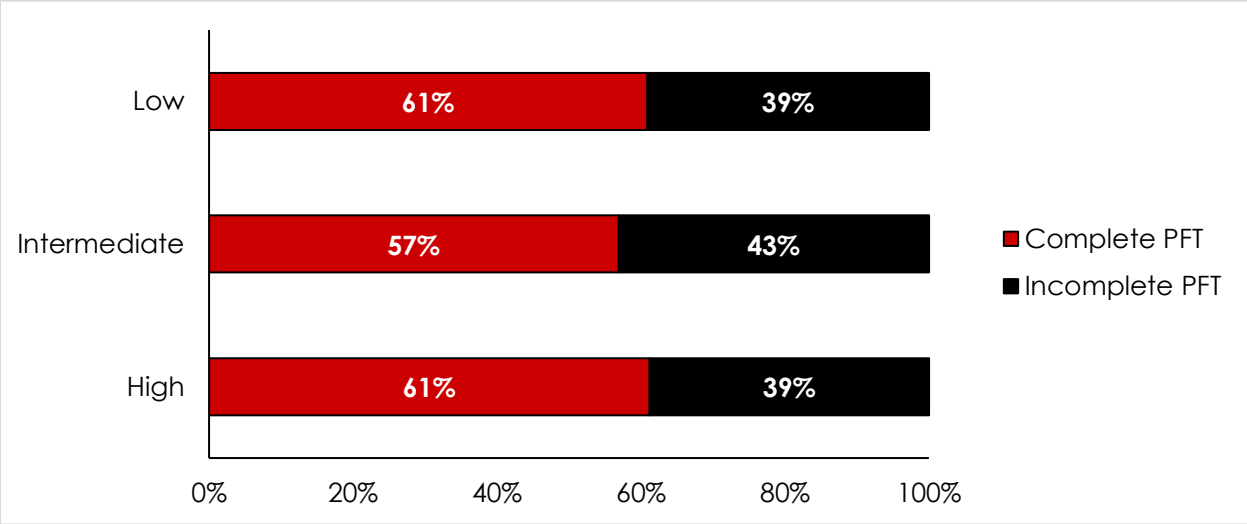
Figure 15: Missing Post-Bronchodilator Values by BMI Classification, Among Non-CCDSS Cases



Frailty

In Figure 16, PFT completeness varies little according to frailty risk (57% to 61%).

Figure 16: Completeness of PFT by Frailty Risk



Figures 17 and 18 describe how the missing airflow data issue varies according to frailty risk, presented according to CCDSS case status. In Figure 17, among CCDSS cases, 15 to 20 percent of individuals are missing airflow data from their PFT records, with the greatest proportion (20%) among individuals in the intermediate-frailty risk group.

A similar trend is evident among non-CCDSS cases (Figure 18), with 14 to 17 percent of individuals missing airflow data and the high-frailty risk group having the greatest proportion (17%) of individuals missing information.

Figure 17: Missing Airflow Data by Frailty Risk, Among CCDSS Cases

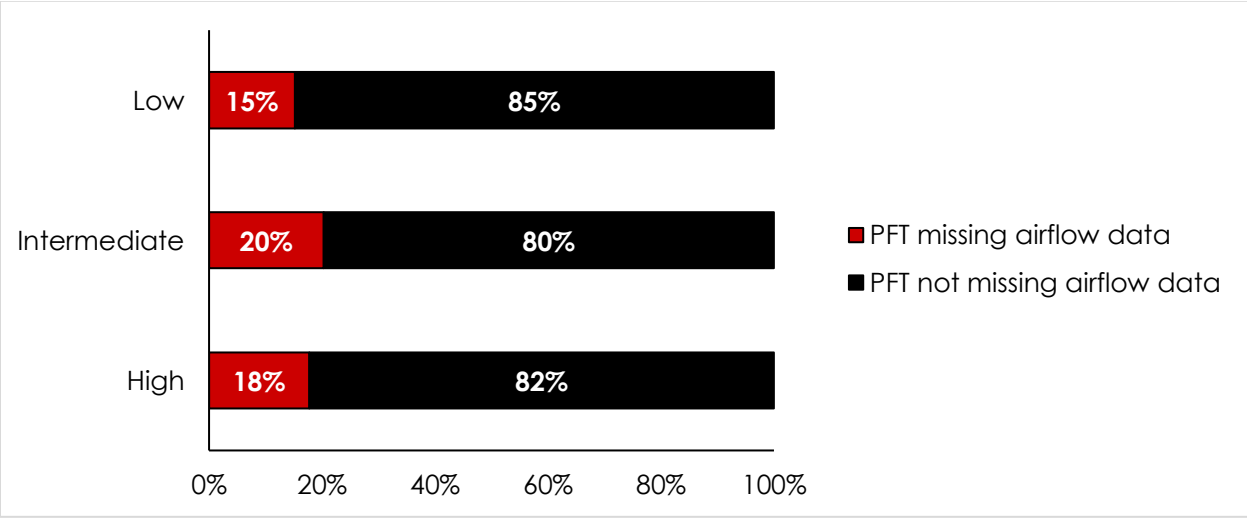
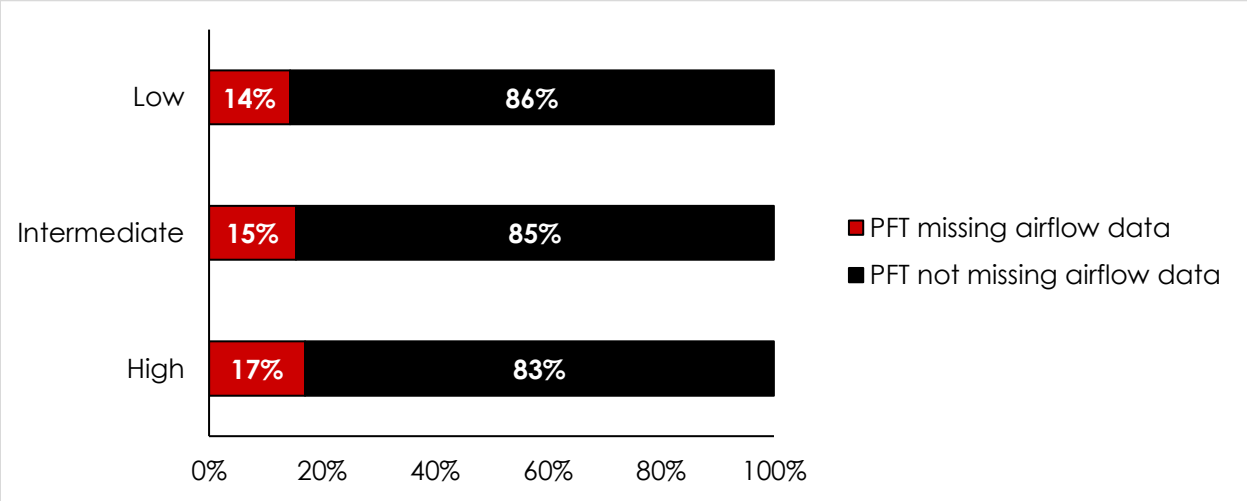


Figure 18: Missing Airflow Data by Frailty Risk, Among Non-CCDSS Cases



Figures 19 and 20 describe how post-bronchodilator application varies according to frailty risk, presented according to CCDSS case status. Among CCDSS cases (Figure 19), the greatest proportion of individuals missing post-bronchodilator values is among individuals with intermediate frailty risk (24%), while both low- and high-frailty risk groups have 20 percent of individuals lacking post-bronchodilator data.

Among non-CCDSS cases (Figure 20), the proportion of individuals without post-bronchodilator test data ranges from 29 to 35 percent, with the intermediate-frailty risk group once again having the largest proportion of individuals missing a post-bronchodilator test.

Figure 19: Missing Post-Bronchodilator Values by Frailty Risk, Among CCDSS Cases

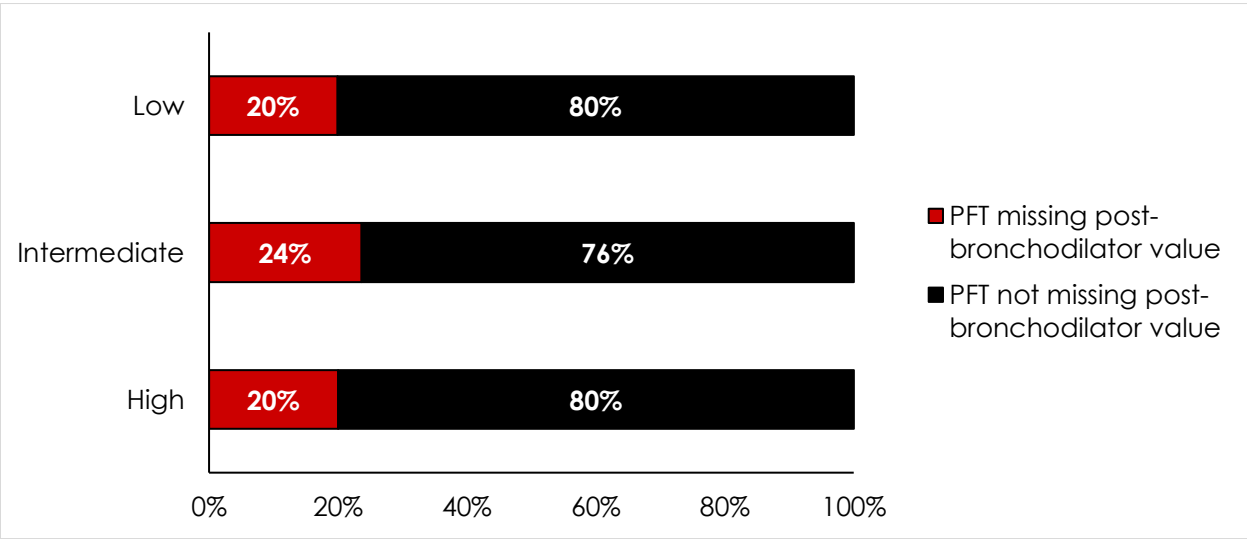
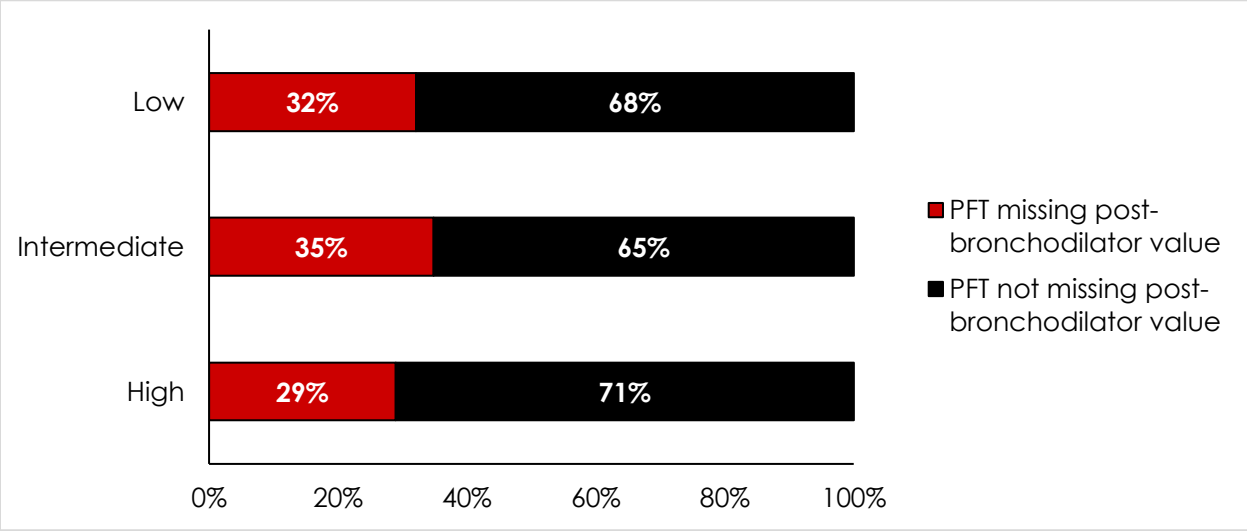


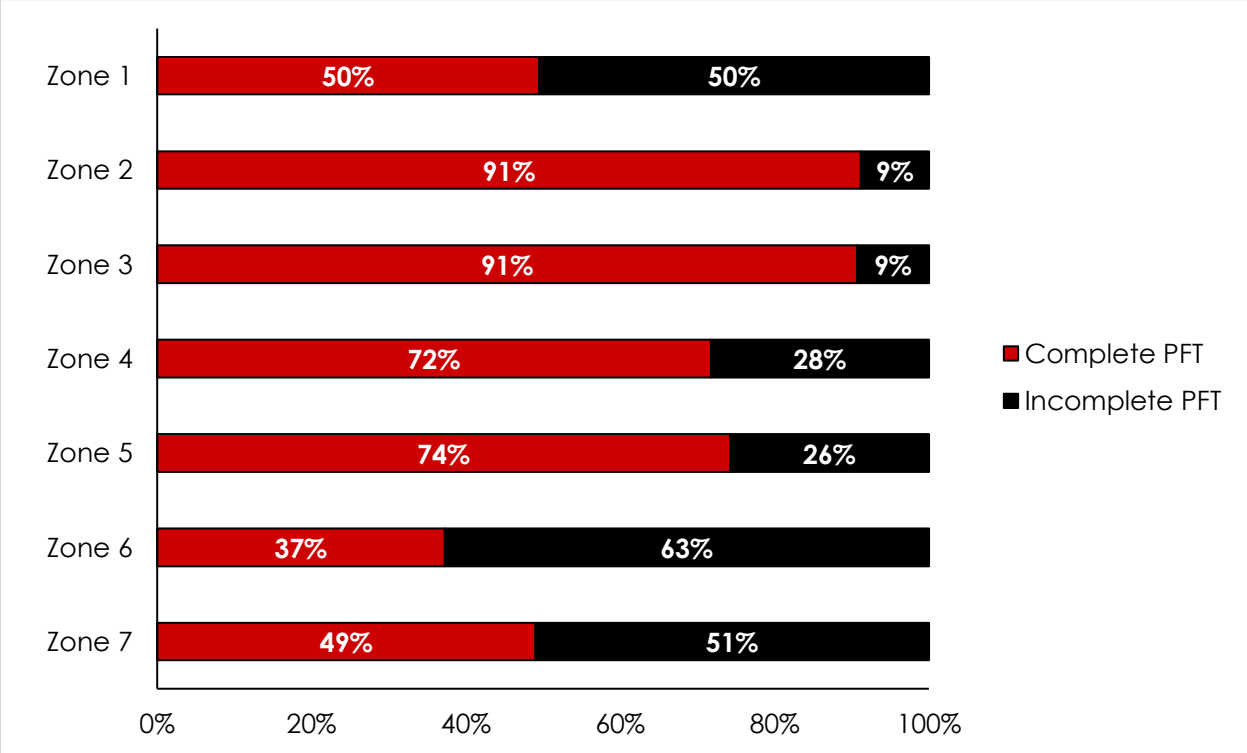
Figure 20: Missing Post-Bronchodilator Values by Frailty Risk, Among Non-CCDSS Cases



Health Zone

Figure 21 describes how PFT completeness varies by health zone. Zones 1 (50%), 6 (63%), and 7 (51%) have at least half of all PFT records missing either airflow data or post-bronchodilator tests, while Zones 4 (28%) and 5 (26%) have just over a quarter of records missing this information. Zones 2 and 3 have the lowest incidence (9%) of PFT records missing information.

Figure 21: Completeness of PFT by Health Zone



Figures 22 to 25 further break down the relationship between health zone and incomplete PFT results described in Figure 21, separately presenting the incidence of missing airflow information and lack of post-bronchodilator testing by health zone and CCDSS case status.

In Figure 22, among CCDSS cases, only two health zones, Zone 1 (27%) and Zone 6 (30%), have a substantial proportion of records missing airflow information. The same pattern is observed in Figure 23 among non-CCDSS cases (Zone 1: 34%; Zone 6: 35%).

Figure 22: Missing Airflow Data by Health Zone, Among CCDSS Cases

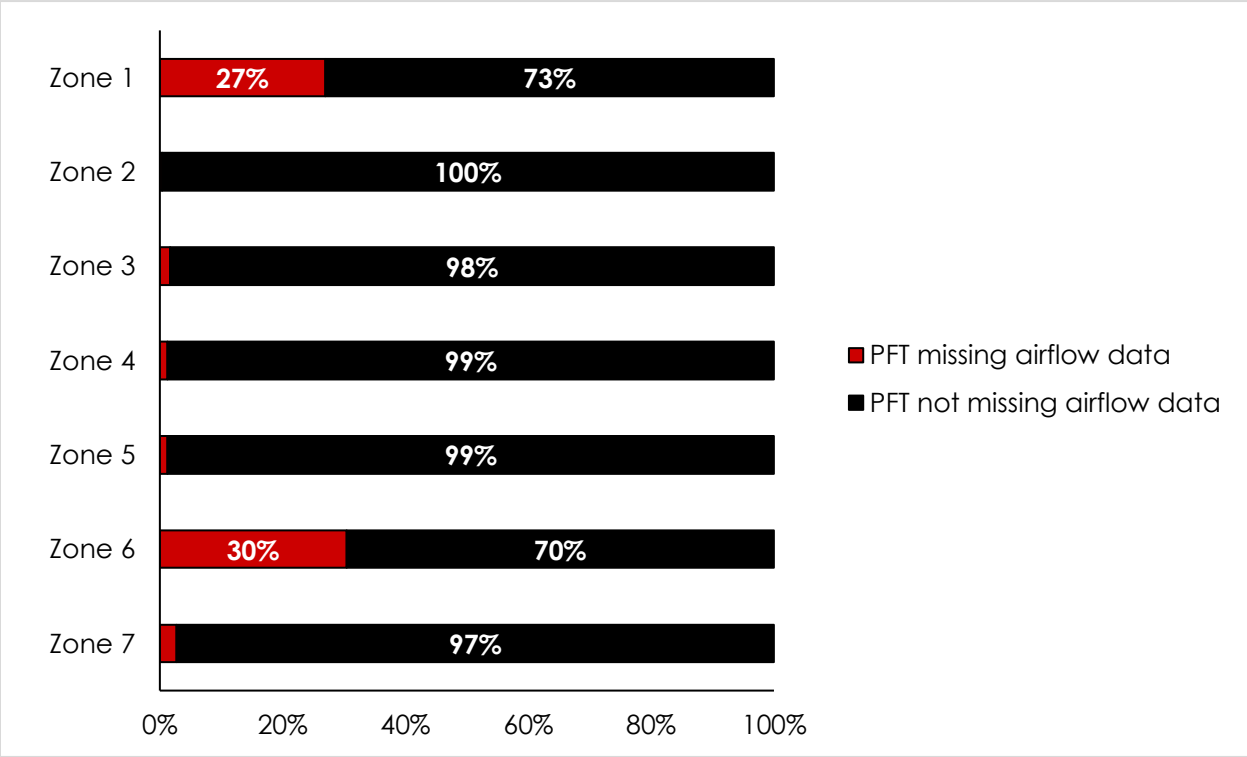


Figure 23: Missing Airflow Data by Health Zone, Among Non-CCDSS Cases

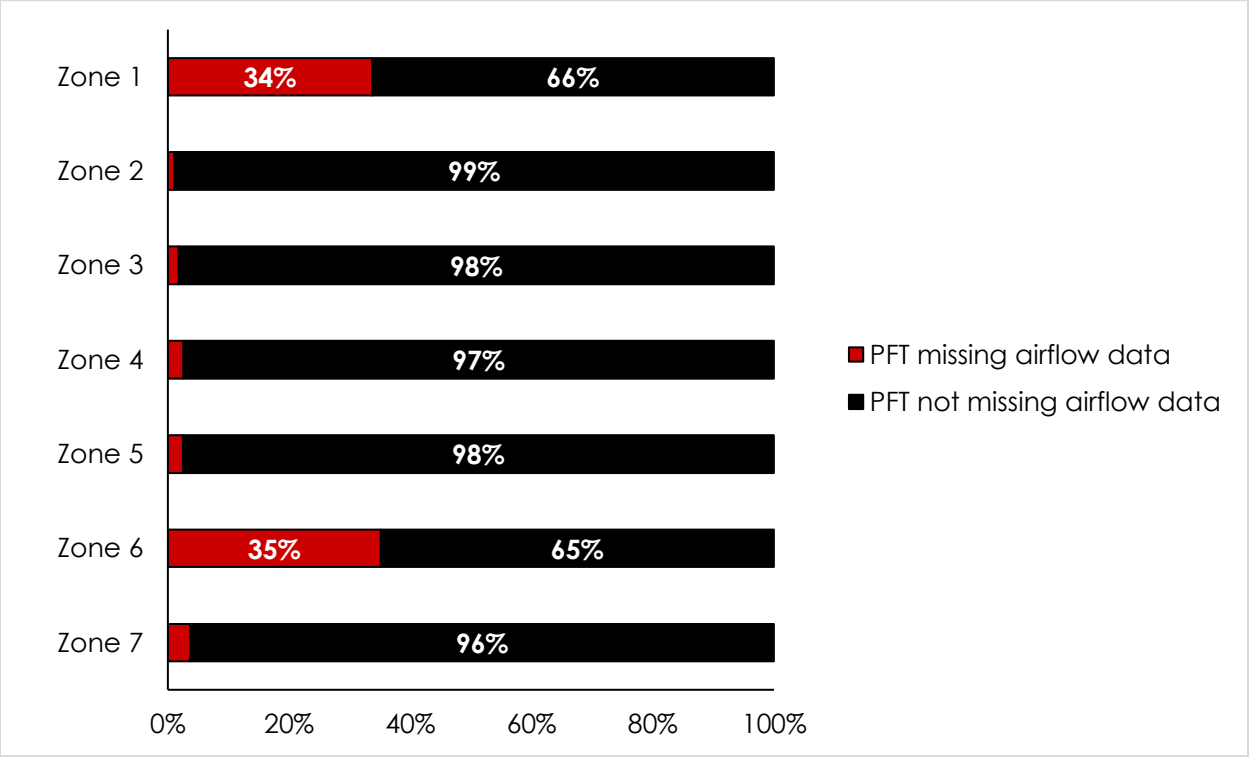


Figure 24, which describes the absence of post-bronchodilator tests by health zone among CCDSS cases, shows that just under one-third of cases from Zones 6 and 7 (32% for both zones) did not have a post-bronchodilator test, while the remaining health zones range from 6 to 17 percent with records lacking post-bronchodilator tests.

A similar pattern of variation across health zones is observed among non-CCDSS cases (Figure 25), with Zones 6 and 7 having the largest proportion of individuals (50% and 54%, respectively) who did not receive post-bronchodilator tests, and proportions ranging from 9 to 33 percent across the remaining health zones.

Figure 24: Missing Post-Bronchodilator Values by Health Zone, Among CCDSS Cases

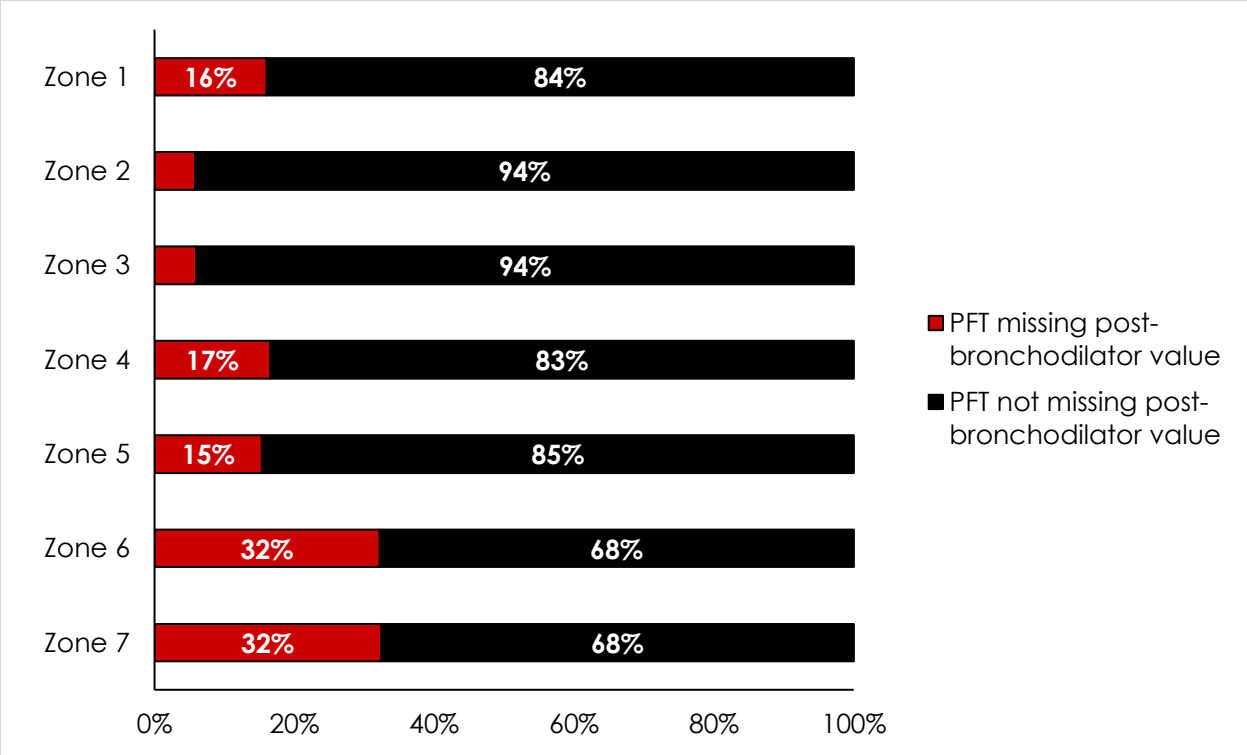
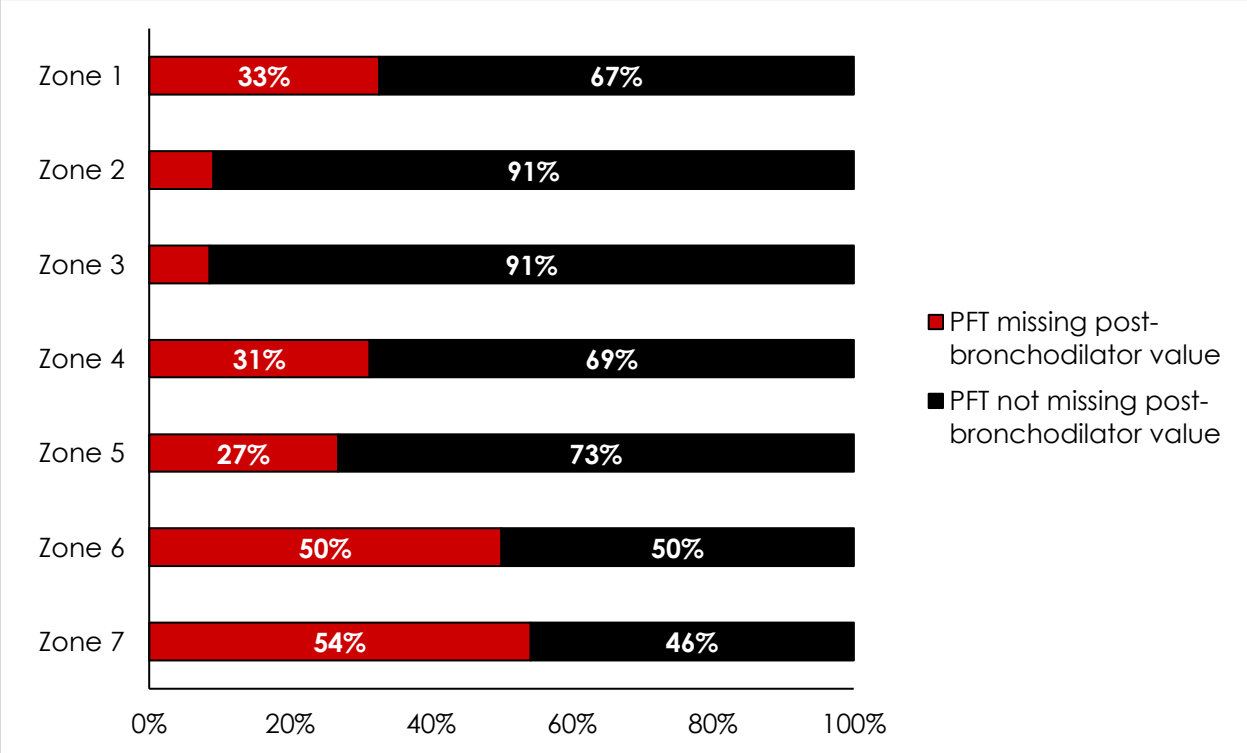


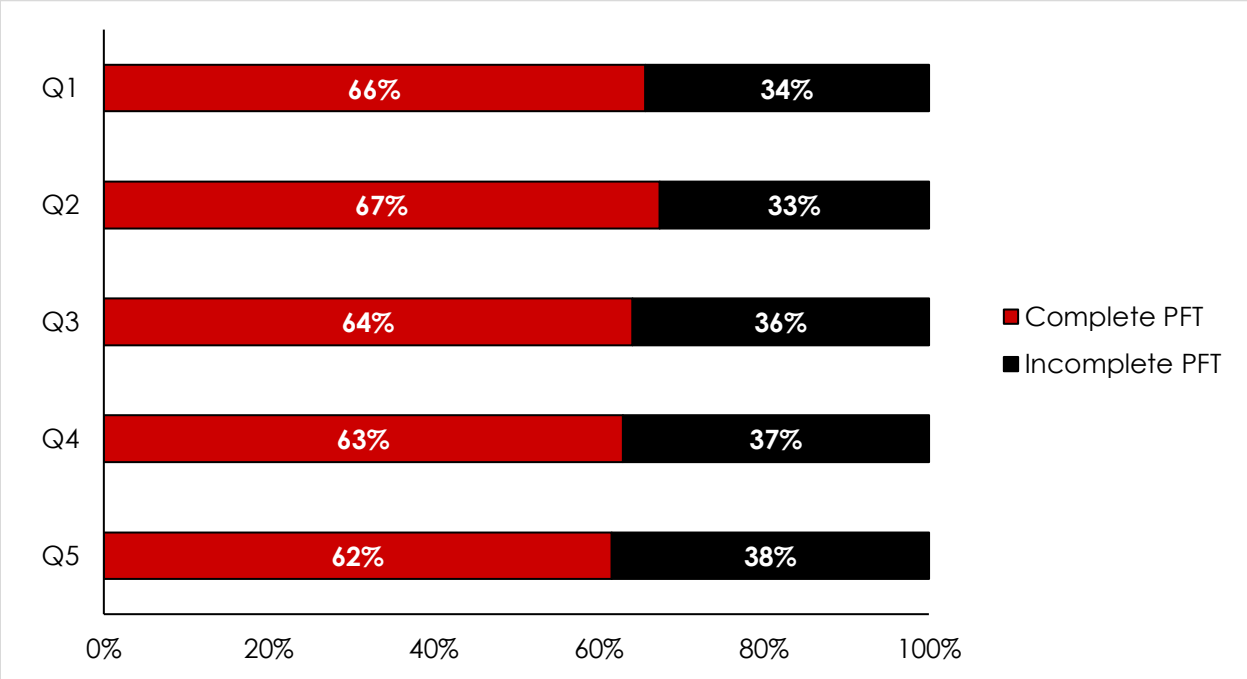
Figure 25: Missing Post-Bronchodilator Values by Health Zone, Among Non-CCDSS Cases



Area Income Quintile

Figure 26 demonstrates a slight gradient in the proportion of incomplete PFTs, which increases from 33 percent (Q2) to 38 percent (Q5) as income increases.

Figure 26: Completeness of PFT by Area Income Quintile



Figures 27 and 28 describe how the missing airflow data issue relates to neighbourhood income, presenting the results by CCDSS case status.

Among CCDSS cases (Figure 27) the proportion of records missing airflow information ranges from 10 percent (Q2) to 17 percent (Q5), with a slight potential trend toward increased proportions of missing data with increasing income.

A similar pattern is observed among non-CCDSS cases (Figure 28), ranging from 13 percent (Q2) to 20 percent (Q5).

Figure 27: Missing Airflow Data by Area Income Quintile, Among CCDSS Cases

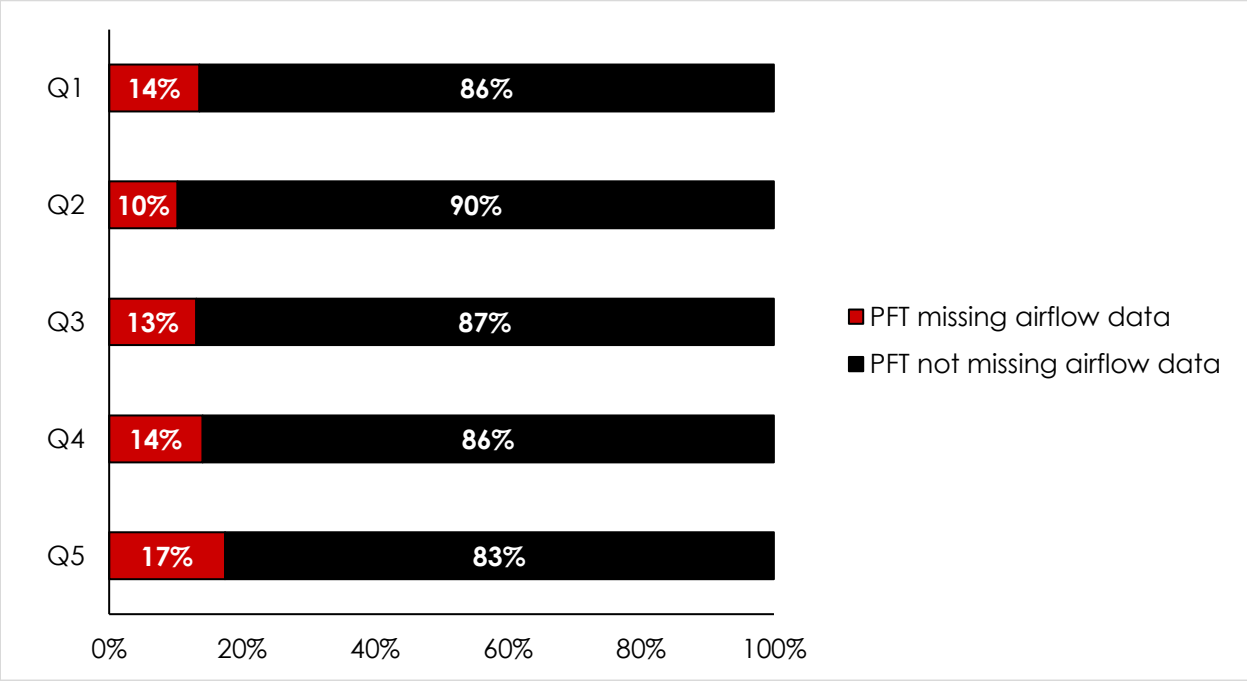
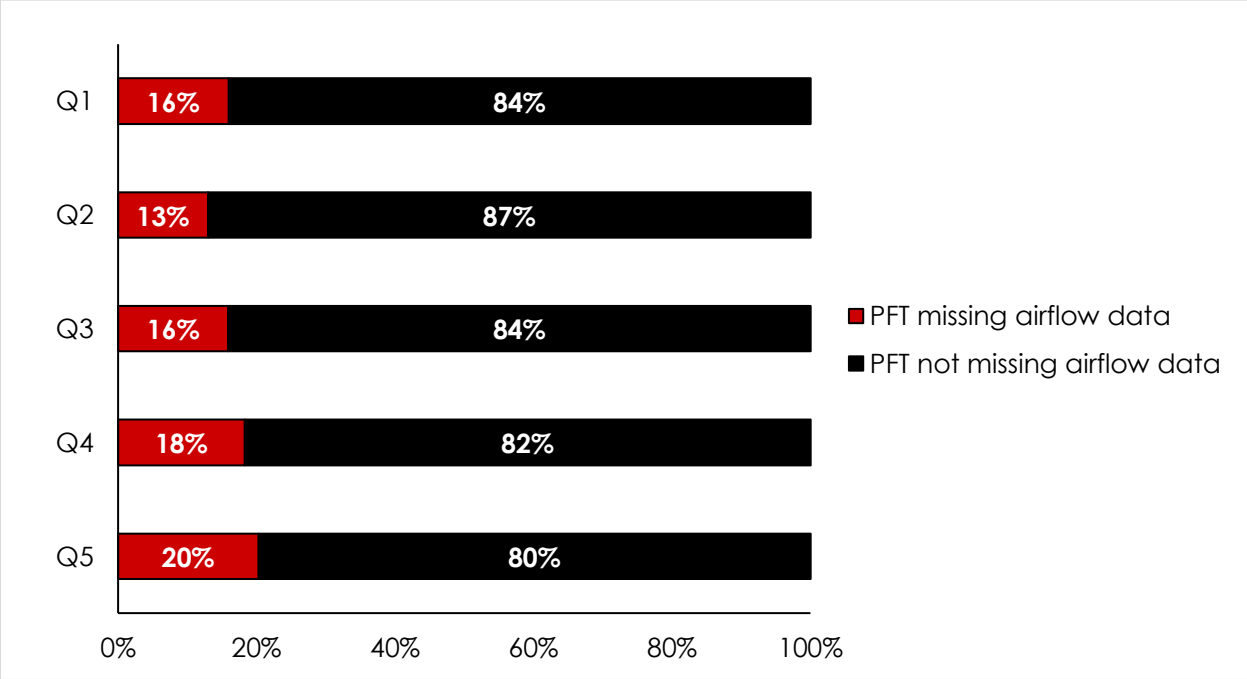


Figure 28: Missing Airflow Data by Area Income Quintile, Among Non-CCDSS Cases



Figures 29 and 30 describe how the lack of post-bronchodilator testing varies according to neighborhood income, presented by CCDSS cases status. Among CCDSS cases, there is little variation in the proportion of individuals missing a post-bronchodilator value, which ranges from

14 to 15 percent across all income quintiles. A similar trend is observed among non-CCDSS cases (Figure 30), with values ranging from 27 to 28 percent.

Figure 29: Missing Post-Bronchodilator Values by Area Income Quintile, Among CCDSS Cases

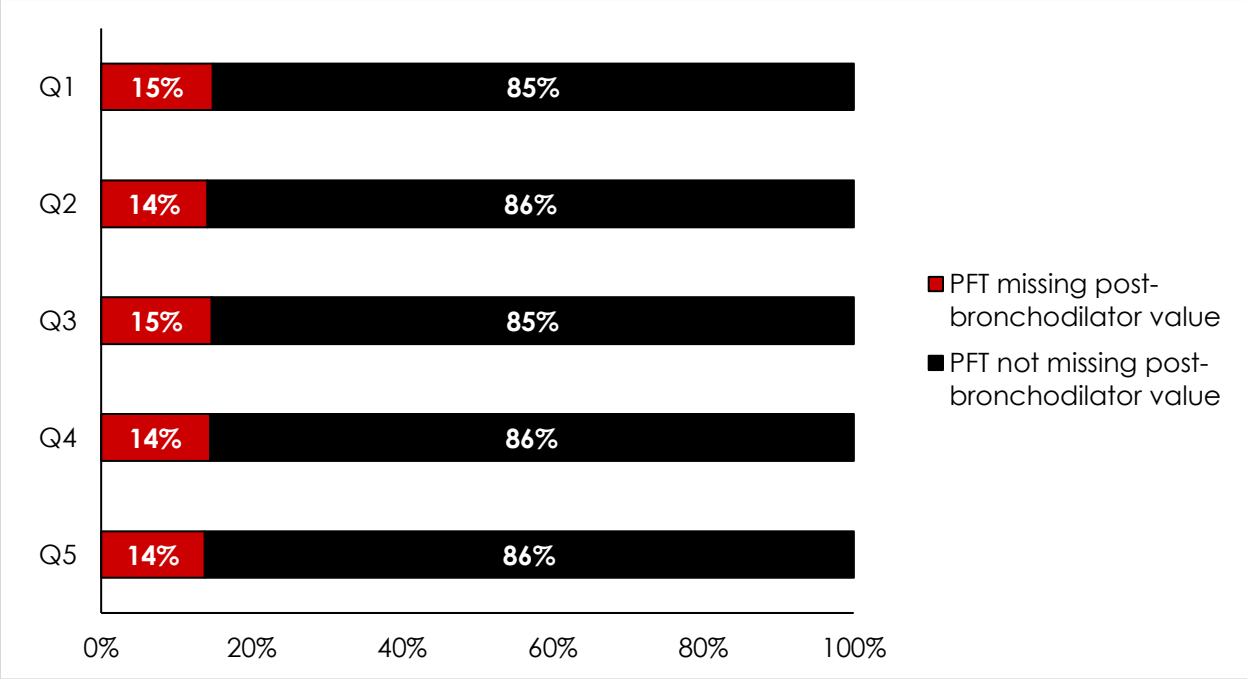
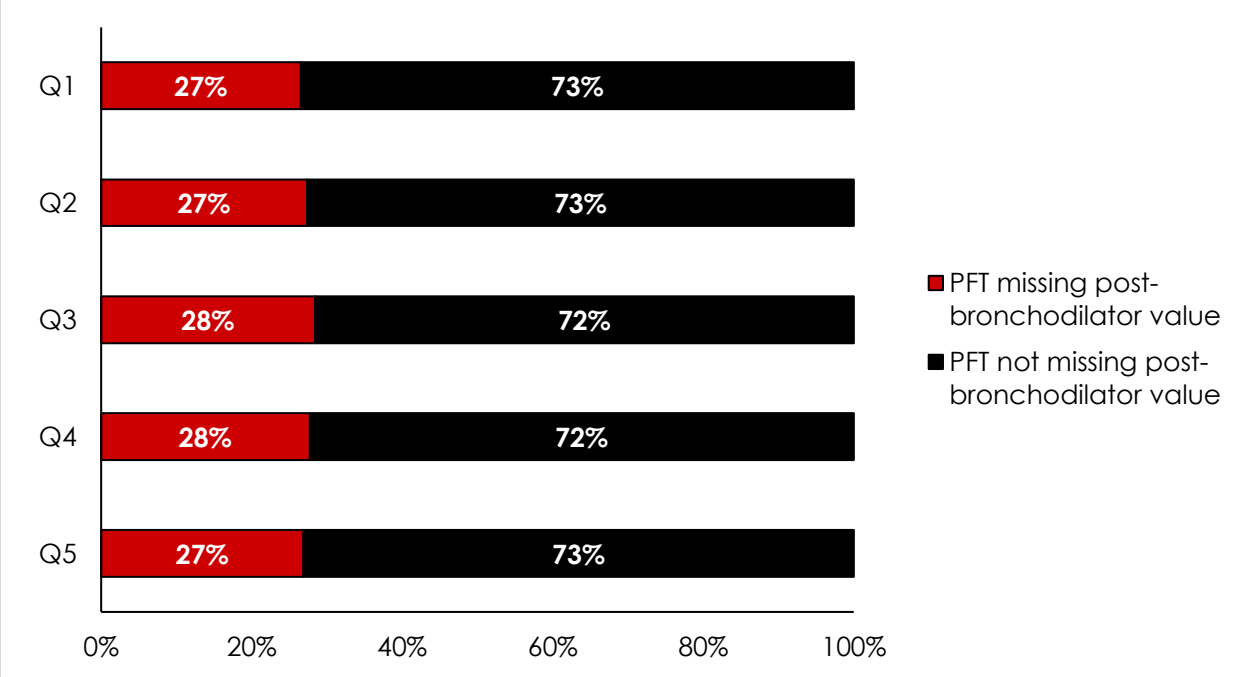


Figure 30: Missing Post-Bronchodilator Values by Area Income Quintile, Among Non-CCDSS Cases



Discussion

When exploring the phenomenon of incomplete PFTs within NB-CHIP, two potential issues were identified.

Data Quality

First, a data quality issue was identified in which a sizable number of records (n=9,050, 10%) were missing all airway measurement data. Currently, the leading explanation is that this missing airway measurement data results from a systemic data quality issue which occurs at the data extraction phase. This explanation is reflected in the results from Figures 22 and 23, which demonstrate that only two zones (each with a single facility reporting to NB-CHIP, both of which were using the same software) had missing airway measurements.

An important consideration is whether these missing data are missing at random, as bias may be introduced into any analyses conducted when data are not missing at random. We observed variation in the proportion of records missing airway measurement data across categories in some of our analyses by characteristic (age, BMI, area income quintile). Such variation may reflect the expected distribution of patient characteristics at the reporting facilities, although it may also be a sign that data are not missing at random. Further analyses are required to disentangle these possibilities.

Absence of Post-Bronchodilator PFT Results

The second issue identified is the absence of post-bronchodilator PFT results, which is likely a clinical issue rather than a data quality issue in most cases. Approximately 13 percent, or 11,505 records, did not have post-bronchodilator values. There are several possible explanations for why post-bronchodilator values may be absent, some of which are valid and expected, and some of which reflect problems that could be improved upon.

Explanation 1: The Suspected Diagnosis Did Not Indicate a Post-Bronchodilator Test

While the absence of a post-bronchodilator PFT is problematic when examining obstructive diseases such as COPD and asthma, post-bronchodilator PFTs are not necessarily indicated for all conditions that could be included in the airway complaints cohort.

NB-CHIP records do not specify the rationale for PFT testing, making it difficult to distinguish cases in which a post-bronchodilator PFT was indicated from cases in which one was not indicated. However, we may assume that a post-bronchodilator PFT was indicated in patients with a COPD diagnosis and may or may not have been indicated in patients without a COPD diagnosis.

When comparing CCDSS COPD cases and non-CCDSS cases, regardless of the explanatory variable used in the analyses, we found that non-CCDSS cases had a greater proportion of

records missing post-bronchodilator results. This observation may be explained, at least in part, by the notion that post-bronchodilator testing is not indicated in some portion of non-CCDSS cases.

However, all analyses showed a proportion of missing post-bronchodilator results even among CCDSS cases (where such testing is presumably indicated), suggesting that appropriate pulmonary function testing is not always performed when indicated. As such, alternative explanations must be explored.

Explanation 2: The Suspected Diagnosis Did Indicate a Post-Bronchodilator Test, but the Test Was Not Performed for a Valid Reason

A second possible explanation for missing post-bronchodilator values is that the care provider opted not to perform the post-bronchodilator test for a valid reason. This could occur when the patient has a contraindication to a bronchodilator or in cases in which the patient is deemed by the provider to be unlikely to be able to successfully complete a post-bronchodilator test (e.g., due to the added effort of back-to-back pre- and post-bronchodilator testing).

This scenario may be reflected in the results which demonstrate that missing post-bronchodilator values are increasingly common with increasing age, as elderly patients may be more likely to have difficulty completing repeated spirometry procedures.

Explanation 3: The Suspected Diagnosis Did Indicate a Post-Bronchodilator Test, but the Test Was Not Performed for an Invalid Reason

A third possible explanation for missing post-bronchodilator values is that the clinician neglected to perform a post-bronchodilator PFT when it was indicated and feasible for the patient. For example, a clinician might neglect to perform a post-bronchodilator test due to their beliefs or practices deviating from recommended guidelines. Unfortunately, however, the available data do not permit exploration of the likelihood or potential extent of this explanation, and none of the disaggregated analyses appear to provide any relevant insights.

Explanation 4: The Suspected Diagnosis Did Indicate a Post-Bronchodilator Test, and the Test Was Performed, but it Either Was Not Recorded or Was Not Captured in the Database Due to Data Processing Problems

A final potential explanation for missing post-bronchodilator values is that a post-bronchodilator test was completed as indicated, but something prevented it from being recorded in the NB-CHIP database. Contributing factors could include deficient documentation practices or technical issues with data collection or data management.

This explanation may be reflected in our analysis of missing post-bronchodilator values disaggregated by health zone. For example, the especially low prevalence of missing values in Zone 2 and Zone 3 may reflect superior documentation and data management practices at facilities located in those zones.

Conclusion

NB-CHIP is a novel data set, and data quality and potential clinical issues have been identified. Paramount among these issues, when considering obstructive diseases such as COPD and asthma, are records missing PFT measurement data necessary for diagnosing and characterizing such diseases. The two sub-issues identified here – records missing all airway information, and records missing a post-bronchodilator PFT result – are likely to have resulted from different underlying causes.

The issue of records missing all airway information, which is a data quality issue, appears to stem from a systematic data extraction error affecting only two facilities. We hope that future software upgrades at these facilities will resolve this issue, and that the missing records can be fully restored and included in future analyses.

The second issue, likely a clinical issue in most cases, may not be resolved as easily due to the nature of the underlying causes. For example, a substantial proportion of missing post-bronchodilator values may be the result of clinical decision making (i.e., the provider not ordering the test), which could include cases in which the test was indicated or not indicated. The former would require changes in clinical practice to address, while the latter is not actually a problem. Deficiencies in reporting and/or technical data processing issues may also account for some of the missing post-bronchodilator values, although further evaluations will be required to understand the extent to which they contribute to the missing values.

This report provides insight into some of the reasons underlying missing data in the NB-CHIP data set. An understanding of the factors contributing to missing data can provide an appreciation of the risks of bias such missing data may introduce, which may assist in the appropriate interpretation of research results.

Furthermore, understanding the sources of missing data may inform strategies to improve data collection and processing practices, in turn improving the overall quality of administrative databases. Despite the data quality issues identified, NB-CHIP, particularly when paired with CCDSS case identification data sets, appears to be a powerful tool for research in New Brunswick.

References

1. New Brunswick Institute for Research, Data and Training. CCDSS Chronic Obstructive Pulmonary Disease (COPD) DH01 c02 codebook for years 1995-2019. Fredericton, NB: New Brunswick Institute for Research, Data and Training; 2021. Available from: https://www.unb.ca/nbirdt/data/holdings/_resources/pdf/dh01c02_2021_dh_ccdss_copd.pdf

2. Government of Canada. Canadian Chronic Disease Surevillance System: Summary of methods: Government of Canada; 2023 Jan 1. Available from: https://health-infobase.canada.ca/ccdss/Methods.

3. New Brunswick Institute for Research, Data and Training. Citizen Database DH10 codebook for years 1971-2021. Fredericton, NB: New Brunswick Institute for Research, Data and Training; 2021. Available from: https://www.unb.ca/nbirdt/data/holdings/_resources/pdf/dh10_2021_dh_citizen_data.pdf

4. New Brunswick Institute for Research, Data and Training. Discharge Abstract Database (DAD) reference guide for years 1999-2014. Fredericton, NB: New Brunswick Institute for Research, Data and Training; 2017. Available from: https://www.unb.ca/nbirdt/_assets/documents/nbirdt/data-holdings/discharge_abstract_database_reference_guide.pdf

5. New Brunswick Institute for Research, Data and Training. NB Chronic Obstructive Pulmonary Disease Health Information Platform (NB-CHIP) - Horizon HH04: variable list. Fredericton, NB: New Brunswick Institute for Research, Data and Training; 2021. Available from: https://www.unb.ca/nbirdt/data/holdings/_resources/pdf/variable-list-hh04-nb-chip-horizon.pdf

6. Global Initiative for Chronic Obstructive Lung Disease, Inc. Global stategy for the diagnosis, management, and prevention of Chronic Obstructive Pulmonary Disease. 2020 [cited 2022 Dec 14]. Available from: https://goldcopd.org/wp-content/uploads/2019/12/GOLD-2020-FINAL-ver1.2-03Dec19_WMV.pdf

7. Statistics Canada. Postal Code OM Conversion File (PCCF), reference guide: November 2014. Ottawa, ON; 2015 Apr 14 [cited 2022 Dec 14]. Statistics Canada Catalogue no. 92-154-G. Available from: https://www150.statcan.gc.ca/n1/en/pub/92-154-g/92-154-g2015001-eng.pdf?st=smYjruM0

8. Government of Canada. Body Mass Index (BMI) Nomogram. 2021 [cited 2022 Dec 14]. Available from: https://www.canada.ca/en/health-canada/services/food-nutrition/healthy-eating/healthy-weights/canadian-guidelines-body-weight-classification-adults/body-mass-index-nomogram.html

9. Gilbert T, Neuburger J, Kraindler J, Keeble E, Smith P, Ariti C, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. Lancet. 2018 [cited 2022 Dec 14];391(10132):1775-82. Available from: https://doi.org/10.1016/S0140-6736(18)30668-8