Validation of diagnostic coding for chronic obstructive pulmonary disease in an electronic health record system in Hong Kong

WC Kwok *, Terence CC Tam, CW Sing, Esther WY Chan, CL Cheung

ABSTRACT

Introduction: Electronic health record databases can facilitate epidemiology research regarding diseases such as chronic obstructive pulmonary disease (COPD), a common medical condition worldwide. We aimed to assess the validity of International Classification of Diseases, 9th Revision (ICD-9) code algorithms for identifying COPD in Hong Kong's territory-wide electronic health record system, the Clinical Data Analysis and Reporting System (CDARS).

Methods: Adult patients diagnosed with COPD at all public hospitals in Hong Kong and specifically at Queen Mary Hospital from 2011 to 2020 were identified using the ICD-9 code 496 (Chronic airway obstruction, not elsewhere classified) within the CDARS. Two respiratory specialists reviewed clinical records and spirometry results to confirm the presence of COPD in a randomly selected group of cases.

Results: During the study period, 93 971 and 2479 patients had the diagnostic code for COPD at all public hospitals in Hong Kong and specifically at Queen Mary Hospital, respectively. Two hundred cases were randomly selected from Queen Mary Hospital for validation using medical records and spirometry results. The overall positive predictive value was 81.5% (95% confidence interval=76.1%-86.9%). We also developed an algorithm to identify COPD cases in our cohort.

Conclusion: This study represents the first validation of ICD-9 coding for COPD in the CDARS. Our findings demonstrated that the ICD-9 code 496 is a reliable indicator for identifying COPD cases, supporting the use of the CDARS database for further clinical research concerning COPD.

Hong Kong Med J 2024;30:380–5
https://doi.org/10.12809/hkmj2210657

[1] **WC Kwok *,** MB, BS, FHKAM (Medicine)
[1] **TCC Tam,** MB, BS, FHKAM (Medicine)
[2] **CW Sing,** PhD
[2] **EWY Chan,** PhD
[2] **CL Cheung,** PhD

[1] *Department of Medicine, Queen Mary Hospital, The University of Hong Kong, Hong Kong SAR, China*
[2] *Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China*

* Corresponding author: kwokwch@hku.hk

This article was published on 29 Aug 2024 at www.hkmj.org.

New knowledge added by this study

- This is the first validation study of International Classification of Diseases, 9th Revision (ICD-9) coding for chronic obstructive pulmonary disease (COPD) in the Hong Kong Clinical Data Analysis and Reporting System (CDARS).
- The ICD-9 code 496 demonstrated a high positive predictive value for identifying COPD cases in the CDARS.

Implications for clinical practice or policy

- This study established an algorithm for identifying COPD cases in the CDARS.
- The findings provide a basis for territory-wide analysis of COPD in Hong Kong.

Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic inflammatory lung disease characterised by airflow limitation, which causes symptoms such as difficulty breathing, productive cough, and wheezing. Smoking is the primary risk factor for COPD development.[1] Patients with COPD experience gradual deterioration of lung function, with potential intermittent exacerbations.

Although COPD is preventable and manageable, it was ranked as the fourth leading cause of death worldwide in the 2019 Global Initiative for Chronic Obstructive Lung Disease guidelines.[2] The Global Burden of Disease Study estimated that there were 3.2 million COPD-related deaths in 2015, an increase of 11.6% compared with 1990.[3] The prevalence of COPD also increased by 44.2% during the same period, reaching 174.5 million cases in 2015.[3] In Hong Kong, the Population Health Survey 2014/15 revealed that 0.5% (0.6% in male individuals; 0.4% in female individuals) of non-institutionalised persons aged ≥15 years had physician-diagnosed COPD.[4]

The prevalence of COPD in Hong Kong

among adults aged ≥60 years is 25.9% or 12.4%, depending on the spirometric criteria used (post-bronchodilator ratio of forced expiratory volume in 1 second to forced vital capacity [ie, FEV_1/FVC ratio] <70% or lower limit of normal).[5] In 2005, the crude mortality rate for COPD was 29.1 per 100 000 population, whereas the crude hospitalisation rate was 193 per 100 000 population.[6] From January 2017 to December 2020, there were 78 693 admissions for COPD across all public hospitals in Hong Kong.[7,8]

Population-based or large database studies are valuable for understanding the epidemiology, clinical characteristics, and burden of COPD.[9-15] In countries/regions with electronic health record (EHR) systems, the EHR databases offer extensive information for clinical management, research, and big data analysis of various diseases, including COPD. Studies in the US and the United Kingdom have validated diagnostic codes for COPD and acute exacerbation of COPD. A study of the diagnostic code for COPD in the US showed a positive predictive value (PPV) of 91.7%, sensitivity of 71.7%, and specificity of 94.4%.[16] In the United Kingdom, the diagnostic code for acute exacerbation of COPD had a PPV of 85.5% and sensitivity of 62.9%.[17] Electronic health records typically contain diagnostic information, associated morbidity and mortality data, and possible longitudinal follow-up data, allowing the evaluation of COPD trends and associated health outcomes. Before research can be conducted using EHR data, the diagnostic coding must be validated. The Clinical Data Analysis and Reporting System (CDARS), an EHR database managed by the Hospital Authority (HA; a public healthcare service provider that manages 43 hospitals/institutions and 123 outpatient clinics[18]), has covered >90% of the Hong Kong population since 1993. The CDARS captures medical information including diagnoses, drug prescriptions, demographics, admissions, medical procedures, and laboratory results. Although the accuracy of diagnostic coding has been demonstrated for some conditions in Hong Kong,[19-21] it has not been validated for COPD. In this study, we aimed to assess the validity of International Classification of Diseases, 9th Revision (ICD-9) code algorithms for identifying COPD in the CDARS.

Methods

This study was conducted at Queen Mary Hospital (QMH), a territory-wide tertiary and quaternary referral centre under HA for advanced medical services and respiratory diseases. All medical information regarding its patients is captured within the CDARS.

Firstly, all adult patients aged ≥40 years with a principal diagnosis of COPD in HA from 1 January 2011 to 31 December 2020 were identified through the CDARS. Then, in the ICD-9 coding validation session, it included adult patients aged ≥40 years with a principal diagnosis of COPD recorded at QMH from 1 January 2011 to 31 December 2020. Potential COPD cases in the CDARS were initially identified using the ICD-9 code 496 (Chronic airway obstruction, not elsewhere classified). Cases with a secondary diagnosis of ICD-9 code 493 (Asthma; indicating potential asthma-COPD overlap [ACO] or asthma) were excluded. The clinical information and spirometry results for all potential COPD cases during the study period were retrieved for validation from the CDARS. The algorithm used for case identification is depicted in Figure 1.

Among potential cases identified in the QMH cohort, 200 were randomly selected for validation. Case validation was performed by two respiratory specialists, based on the clinical information, spirometry results, physician notes, and clinical examination reports. A potential COPD case was regarded as true positive if the specialist concluded that the patient had definite COPD according to the Global Initiative for Chronic Obstructive Lung Disease guidelines.[22] A valid case was defined as the presence of symptoms compatible with COPD, along with spirometry results demonstrating airflow limitation (ie, FEV_1/FVC ratio <0.7) that could not be fully reversed by the administration of an inhaled bronchodilator. Potential cases not meeting these criteria were regarded as false positive. Patients without spirometry data were excluded from the

香港電子健康記錄系統中慢性阻塞性肺病診斷編碼的驗證

郭宏駿、譚子雋、成楚頴、陳慧賢、張正龍

引言：電子健康記錄資料庫可以促進慢性阻塞性肺病等全球常見疾病的流行病學研究。我們的目的是評估國際疾病分類第九版修訂版（ICD-9）代碼演算法在香港電子健康記錄系統（亦即臨床數據分析及報告）中識別慢性阻塞性肺病的有效性。

方法：我們使用臨床數據分析及報告內的ICD-9代碼496（慢性氣道阻塞，未分類）識別2011至2020年在香港所有公立醫院，特別是瑪麗醫院診斷患有慢性阻塞性肺病的成年患者。兩位呼吸系統專科醫生審查了臨床記錄和肺功能檢查結果數據，以確認隨機選擇的一組病例中是否存在慢性阻塞性肺病。

結果：研究期間，香港所有公立醫院（特別是瑪麗醫院）分別有93 971名和2479名患者有慢性阻塞性肺病的診斷代碼。我們隨機選擇了瑪麗醫院的200個病例，利用醫療記錄和肺功能檢查結果進行驗證。整體陽性預測值為81.5%（95%置信區間=76.1%-86.9%）。我們還開發了一種演算法來識別隊列中的慢性阻塞性肺病病例。

結論：本研究首次在臨床數據分析及報告中驗證了阻塞性肺病的ICD-9編碼。我們的研究結果表明ICD-9代碼496是識別慢性阻塞性肺病病例的可靠指標，支持使用臨床數據分析及報告資料庫進行有關阻塞性肺病的進一步臨床研究。

case validation process. The flow of patient selection is illustrated in Figure 1.

The PPV was computed to assess the validity of COPD diagnostic codes in the CDARS, using the definition of the number of true positives (ie, cases identified by ICD-9 codes which met the above criteria) divided by the total number of true positives plus false positives (ie, cases identified by ICD-9 codes which did not meet the above criteria).

$$PPV = \frac{Number\ of\ true\ positives}{Number\ of\ true\ positives + false\ positives}$$

Cohen's kappa was used to estimate inter-rater reliability and the 95% confidence interval was estimated using a binomial distribution. All statistical analyses were performed using SPSS software (Windows 26.0; IBM Corp, Armonk [NY], US).

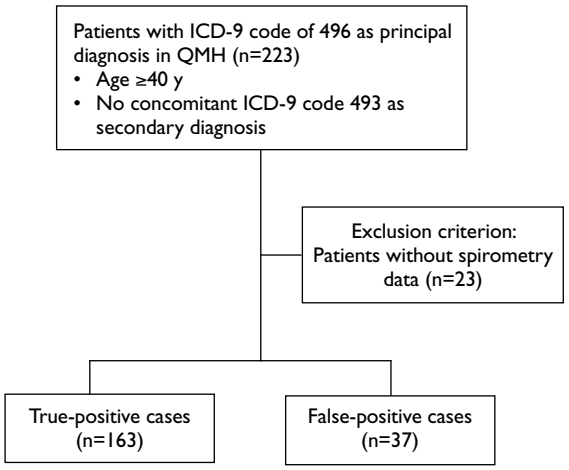

FIG 1. Algorithm for identifying chronic obstructive pulmonary disease cases in the Clinical Data Analysis and Reporting System

Abbreviations: ICD-9 = International Classification of Diseases, 9th Revision; QMH = Queen Mary Hospital

Results

In total, 2479 potential cases were identified in QMH between 2011 and 2020. During the same period, there were 93 971 cases with a principal diagnostic code of COPD across all public hospitals in Hong Kong. There were no significant differences in age or sex between QMH cases and overall cases throughout the HA (Table 1). Of the QMH cases, 200 were randomly selected for detailed validation. The validation process showed that 163 cases were true positives, resulting in an overall PPV of 81.5% (95% confidence interval=76.1%-86.9%). Major reasons for false positives included ACO, asthma, and bronchiectasis (Table 2). Cohen's kappa was 0.77, suggesting substantial agreement. The proposed algorithm for identifying COPD cases in the CDARS is illustrated in Figure 2.

Discussion

In this validation study, the estimated overall PPV was 81.5% when ICD-9 coding was used to identify COPD cases within the CDARS, the territory-wide EHR system in Hong Kong.

A PubMed search using the terms 'COPD' AND 'validation' OR 'international classification of disease codes' did not identify any literature regarding validation of diagnostic codes for COPD in EHRs within Hong Kong. Validation of local diagnostic codes for COPD will facilitate large-scale studies in Hong Kong, which are needed considering the high local prevalence of this disease. Our study showed a PPV >70%, which is the typical validation criterion for case-finding algorithms in population-based cohort studies.[23,24] The high PPV in our study may be attributable to the nature of the CDARS database, with high PPV also reported in other local validation studies involving other diseases.[21,25] The CDARS database contains EHRs from all public hospitals, where diagnostic facilities and diagnostic protocols are well-established; in contrast, data from claims

TABLE 1. Patient characteristics in all chronic obstructive pulmonary disease cases, 2011-2020

	All COPD cases in HA	All COPD cases in QMH	Validated COPD cases in QMH	True-positive cases in validation cohort	P value
No. of cases in CDARS	93 971	2479	200	163	
Age, y	75.8 ± 10.3	76.5 ± 10.8	78.1 ± 10.7	79.5 ± 10.6	0.06
Sex					0.119
Male	76 774 (81.7%)	2004 (80.8%)	172 (86.0%)	140 (85.9%)	
Female	17 197 (18.3%)	475 (19.2%)	28 (14.0%)	23 (14.1%)	

Abbreviations: CDARS = Clinical Data Analysis and Reporting System; COPD = chronic obstructive pulmonary disease; HA = Hospital Authority; QMH = Queen Mary Hospital

* Data are shown as No. (%) or mean±standard deviation, unless otherwise specified

TABLE 2. Reasons for false-positive cases (n=200)*

Asthma-COPD overlap	18 (9.0%)
Bronchiectasis	5 (2.5%)
Asthma	4 (2.0%)
Heart failure	3 (1.5%)
Pneumonia	3 (1.5%)
Interstitial lung diseases	2 (1.0%)
Obstructive sleep apnoea	1 (0.5%)
Silicosis	1 (0.5%)

Abbreviation: COPD = chronic obstructive pulmonary disease
* Data are shown as No. (%)

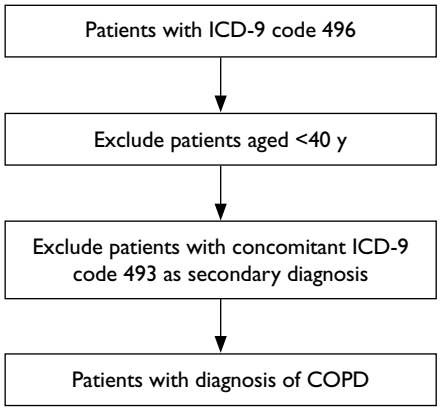

FIG 2. Chronic obstructive pulmonary disease validation
Abbreviations: COPD = chronic obstructive pulmonary disease; ICD-9 = International Classification of Diseases, 9th Revision

databases and general practitioners are expected to have lower accuracy. As such, in prior local validation studies with CDARS, they had high reported PPV of 79%[25] and 100%[21] for interstitial lung diseases and hip fracture, respectively. Also, COPD is a disease that is easier to be recognised by demonstrating airflow obstruction on spirometry, which contributed to the high PPV. Additionally, regular audits by the HA of diagnostic codes in patient discharge summaries to make sure the correct diagnosis were entered further enhance the accuracy of CDARS data.

Among the false-positive cases, ACO was the most frequent cause (Table 2). This relationship could be due to incorrect entry of COPD diagnostic codes or to patients with childhood asthma who developed COPD later in life. The lack of a separate ICD-9 diagnostic code for ACO and the absence of diagnostic criteria for this condition contribute to these challenges.[26-34] Considering the current difficulties in accurate diagnosis of ACO, the actual PPV for COPD could be higher. Thus, our proposed algorithm excludes cases with a secondary diagnosis of asthma in the CDARS to avoid including patients with ACO. Proper education to address this miscoding issue is essential. Asthma was the second most common incorrectly coded diagnosis. This result could be related to initial misdiagnosis at presentation, such as attributing shortness of breath in a smoker to COPD, rather than asthma. Heart failure, which also presents with dyspnoea and wheezing, could be misclassified as COPD in rare instances. Bronchiectasis, pneumonia, silicosis, and interstitial lung disease can also present with chronic productive cough and dyspnoea, similar to COPD.

Strengths and limitations

The strengths of this study include its use of territory-wide database with >11 million records, which allowed the identification of a sufficient number of cases. The methodology utilised to confirm true-positive COPD cases was both feasible and practical: the medical records and spirometry results for all cases with the COPD diagnostic code were reviewed by respiratory specialists.

However, this study had some limitations. First, the patient population mostly comprised adult Chinese patients, consistent with the demographics of patients with COPD in Hong Kong. This ethnicity component may limit generalisability to other populations. Second, only QMH cases were selected for validation. However, because all hospitals and clinics within the HA use a single diagnostic coding system, the diagnostic coding consistency is expected to be high. The high accuracy of ICD-9 coding within the Hong Kong CDARS has been demonstrated in other studies.[20,21]

Conclusion

This study represents the first validation of ICD-9 coding for COPD in Hong Kong. Our findings demonstrated that use of ICD-9 code 496, in conjunction with our algorithm to identify COPD, results in a PPV with sufficient reliability to support utilisation of the CDARS database for future COPD research.

Author contributions

Concept or design: WC Kwok, CL Cheung.
Acquisition of data: WC Kwok.
Analysis or interpretation of data: WC Kwok.
Drafting of the manuscript: WC Kwok, CL Cheung.
Critical revision of the manuscript for important intellectual content: TCC Tam, CW Sing, EWY Chan, CL Cheung.

All authors had full access to the data, contributed to the study, approved the final version for publication, and take responsibility for its accuracy and integrity.

Conflicts of interest

All authors have disclosed no conflicts of interest.

Funding/support

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

Ethics approval

The research was approved by the Institutional Review Board of The University of Hong Kong / Hospital Authority Hong Kong West Cluster, Hong Kong (Ref No.: UW22-716). The requirement for informed consent is waived by the Board due to the retrospective nature of the research.

References

1. Gershon AS, Warner L, Cascagnette P, Victor JC, To T. Lifetime risk of developing chronic obstructive pulmonary disease: a longitudinal population study. Lancet 2011;378:991-6.
2. Singh D, Agusti A, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease: the GOLD Science Committee Report 2019. Eur Respir J 2019;53:1900164.
3. Benziger CP, Roth GA, Moran AE. The Global Burden of Disease study and the preventable burden of NCD. Glob Heart 2016;11:393-7.
4. Centre for Health Protection, Department of Health, Hong Kong SAR Government. Non-Communicable Diseases Watch. Chronic obstructive pulmonary disease: an overview. November 2018. Available from: https://www.chp.gov.hk/files/pdf/ncd_watch_november_2018.pdf. Accessed 9 Aug 2024.
5. Chan HS, Ko FW, Chan JW, et al. Hospital burden of chronic obstructive pulmonary disease in Hong Kong—the trend from 2006 to 2014. Int J Chron Obstruct Pulmon Dis 2023;18:507-19.
6. Chan-Yeung M, Lai CK, Chan KS, et al. The burden of lung disease in Hong Kong: a report from the Hong Kong Thoracic Society. Respirology 2008;13 Suppl 4:S133-65.
7. Ko FW, Lau LH, Ng SS, et al. Respiratory admissions before and during the COVID-19 pandemic with mediation analysis of air pollutants, mask-wearing and influenza rates. Respirology 2023;28:47-55.
8. Chan KP, Ma TF, Kwok WC, et al. Significant reduction in hospital admissions for acute exacerbation of chronic obstructive pulmonary disease in Hong Kong during coronavirus disease 2019 pandemic. Respir Med 2020;171:106085.
9. Buist AS, McBurnie MA, Vollmer WM, et al. International variation in the prevalence of COPD (the BOLD study): a population-based prevalence study. Lancet 2007;370:741-50.
10. Grahn K, Gustavsson P, Andersson T, et al. Occupational exposure to particles and increased risk of developing chronic obstructive pulmonary disease (COPD): a population-based cohort study in Stockholm, Sweden. Environ Res 2021;200:111739.
11. Lindberg A, Lindberg L, Sawalha S, et al. Large underreporting of COPD as cause of death-results from a population-based cohort study. Respir Med 2021;186:106518.
12. Landt E, Çolak Y, Lange P, Laursen LC, Nordestgaard BG, Dahl M. Chronic cough in individuals with COPD: a population-based cohort study. Chest 2020;157:1446-54.
13. Lee SC, Son KJ, Han CH, Park SC, Jung JY. Impact of COPD on COVID-19 prognosis: a nationwide population-based study in South Korea. Sci Rep 2021;11:3735.
14. Du Y, Li Q, Sidorenkov G, et al. Computed tomography screening for early lung cancer, COPD and cardiovascular disease in Shanghai: rationale and design of a population-based comparative study. Acad Radiol 2021;28:36-45.
15. Bahremand T, Etminan M, Roshan-Moniri N, De Vera MA, Tavakoli H, Sadatsafavi M. Are COPD prescription patterns aligned with guidelines? Evidence from a Canadian population-based study. Int J Chron Obstruct Pulmon Dis 2021;16:751-9.
16. Chu SH, Wan ES, Cho MH, et al. An independently validated, portable algorithm for the rapid identification of COPD patients using electronic health records. Sci Rep 2021;11:19959.
17. Rothnie KJ, Müllerová H, Hurst JR, et al. Validation of the recording of acute exacerbations of COPD in UK primary care electronic healthcare records. PLoS One 2016;11:e0151357.
18. Hospital Authority, Hong Kong SAR Government. Introduction. Caring for our community's health. 2024. Available from: https://www.ha.org.hk/visitor/ha_visitor_index.asp?Content_ID=10008&Lang=ENG&Dimension=100&Parent_ID=10004. Accessed 9 Aug 2024.
19. Chan SM, Chung GK, Chan YH, et al. Resilience and coping strategies of older adults in Hong Kong during COVID-19 pandemic: a mixed methods study. BMC Geriatr 2022;22:299.
20. Cheung CL, Tan KC, Kung AW. Cohort profile: the Hong Kong Osteoporosis study and the follow-up study. Int J Epidemiol 2018;47:397-8f.
21. Sing CW, Woo YC, Lee AC, et al. Validity of major osteoporotic fracture diagnosis codes in the Clinical Data Analysis and Reporting System in Hong Kong. Pharmacoepidemiol Drug Saf 2017;26:973-6.
22. Global Initiative for Chronic Obstructive Lung Disease. 2022 Global Strategy for Prevention, Diagnosis and Management of COPD.
23. Cho SK, Doyle TJ, Lee H, et al. Validation of claims-based algorithms to identify interstitial lung disease in patients with rheumatoid arthritis. Semin Arthritis Rheum 2020;50:592-7.
24. Papani R, Sharma G, Agarwal A, et al. Validation of claims-based algorithms for pulmonary arterial hypertension. Pulm Circ 2018;8:2045894018759246.
25. Ye Y, Hubbard R, Li GH, et al. Validation of diagnostic coding for interstitial lung diseases in an electronic health record system in Hong Kong. Pharmacoepidemiol Drug Saf 2022;31:519-23.
26. Leung JM, Sin DD. Asthma-COPD overlap syndrome: pathogenesis, clinical features, and therapeutic targets. BMJ 2017;358:j3772.
27. Cosio BG, Soriano JB, López-Campos JL, et al. Defining the asthma-COPD overlap syndrome in a COPD cohort. Chest 2016;149:45-52.
28. Gibson PG, Simpson JL. The overlap syndrome of asthma and COPD: what are its features and how important is it? Thorax 2009;64:728-35.
29. Sin DD, Miravitlles M, Mannino DM, et al. What is asthma-COPD overlap syndrome? Towards a consensus definition

from a round table discussion. Eur Respir J 2016;48:664-73.
30. Cataldo D, Corhay JL, Derom E, et al. A Belgian survey on the diagnosis of asthma-COPD overlap syndrome. Int J Chron Obstruct Pulmon Dis 2017;12:601-13.
31. Soler-Cataluña JJ, Cosío B, Izquierdo JL, et al. Consensus document on the overlap phenotype COPD-asthma in COPD [in English, Spanish]. Arch Bronconeumol 2012;48:331-7.
32. Koblizek V, Chlumsky J, Zindr V, et al. Chronic obstructive pulmonary disease: official diagnosis and treatment guidelines of the Czech Pneumological and Phthisiological Society; a novel phenotypic approach to COPD with patient-oriented care. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub 2013;157:189-201.
33. Miravitlles M, Alvarez-Gutierrez FJ, Calle M, et al. Algorithm for identification of asthma-COPD overlap: consensus between the Spanish COPD and asthma guidelines. Eur Respir J 2017;49:1700068.
34. Global Initiative for Asthma; Global Initiative for Chronic Obstructive Lung Disease. Diagnosis of Diseases of Chronic Airflow Limitation: Asthma, COPD and Asthma-COPD Overlap Syndrome. Updated 2015. Available from: https://goldcopd.org/wp-content/uploads/2016/04/GOLD_ACOS_2015.pdf. Accessed 2 Aug 2024.