




# Validation of ICD-10-CM Diagnostic Codes for Identifying Patients with ST-Elevation and Non-ST-Elevation Myocardial Infarction in a National Health Insurance Claims Database

Tou-Yuan Tsai <sup>1-3,\*</sup>, Jen-Feng Lin<sup>4,\*</sup>, Yu-Kang Tu<sup>3</sup>, Jian-Heng Lee<sup>4</sup>, Yu-Ting Hsiao<sup>4</sup>, Sheng-Feng Sung <sup>5</sup>, Ming-Jen Tsai <sup>4</sup>

<sup>1</sup>Department of Emergency Medicine, Dalin Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Chiayi, Taiwan; <sup>2</sup>School of Medicine, Tzu Chi University, Hualien, Taiwan; <sup>3</sup>Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan; <sup>4</sup>Department of Emergency Medicine, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan; <sup>5</sup>Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan

\*These authors contributed equally to this work

Correspondence: Sheng-Feng Sung, Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan, Tel +886 5 276 5041 Ext 7283, Fax +886 5 278 4257, Email [sfusng@cych.org.tw](mailto:sfusng@cych.org.tw); [richard.sfsung@gmail.com](mailto:richard.sfsung@gmail.com); Ming-Jen Tsai, Department of Emergency Medicine, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan, Tel +886 5 276 5041 Ext 1909, Fax +886 5 277 4511, Email [tshi33@gmail.com](mailto:tshi33@gmail.com)

**Purpose:** Distinguishing ST-elevation myocardial infarction (STEMI) and non-ST-elevation myocardial infarction (NSTEMI) is crucial in acute myocardial infarction (AMI) research due to their distinct characteristics. However, the accuracy of International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes for STEMI and NSTEMI in Taiwan's National Health Insurance (NHI) database remains unvalidated. Therefore, we developed and validated case definition algorithms for STEMI and NSTEMI using ICD-10-CM and NHI billing codes.

**Patients and Methods:** We obtained claims data and medical records of inpatient visits from 2016 to 2021 from the hospital's research-based database. Potential STEMI and NSTEMI cases were identified using diagnostic codes, keywords, and procedure codes associated with AMI. Chart reviews were then conducted to confirm the cases. The performance of the developed algorithms for STEMI and NSTEMI was assessed and subsequently externally validated.

**Results:** The algorithm that defined STEMI as any STEMI ICD code in the first three diagnosis fields had the highest performance, with a sensitivity of 93.6% (95% confidence interval [CI], 91.7–95.2%), a positive predictive value (PPV) of 89.4% (95% CI, 87.1–91.4%), and a kappa of 0.914 (95% CI, 0.900–0.928). The algorithm that used the NSTEMI ICD code listed in any diagnosis field performed best in identifying NSTEMI, with a sensitivity of 82.6% (95% CI, 80.7–84.4%), a PPV of 96.5% (95% CI, 95.4–97.4%), and a kappa of 0.889 (95% CI, 0.878–0.901). The algorithm that included either STEMI or NSTEMI ICD codes listed in any diagnosis field showed excellent performance in defining AMI, with a sensitivity of 89.4% (95% CI, 88.2–90.6%), a PPV of 95.6% (95% CI, 94.7–96.4%), and a kappa of 0.923 (95% CI, 0.915–0.931). External validation confirmed these algorithms' efficacy.

**Conclusion:** Our results provide valuable reference algorithms for identifying STEMI and NSTEMI cases in Taiwan's NHI database.

**Keywords:** administrative claims data, acute myocardial infarction, diagnosis, ICD-10-CM, non-ST-elevation myocardial infarction, ST-elevation myocardial infarction

## Introduction

Acute myocardial infarction (AMI) is a serious medical condition characterized by myocardial injury or death due to a sudden reduction in blood supply to the heart. Despite significant advances in AMI prognosis in the last decade, it remains a leading cause of morbidity and mortality worldwide.<sup>1-3</sup> AMI is traditionally classified into two types: ST-elevation

myocardial infarction (STEMI) and non-ST-elevation myocardial infarction (NSTEMI), based on the manifestation observed on the electrocardiogram (ECG). Although STEMI and NSTEMI share similar risk factors and etiology, they differ in terms of pathophysiology, epidemiology, and severity.<sup>1–4</sup> Therefore, the European Society of Cardiology (ESC) and the American Heart Association (AHA) regularly update guidelines for the management of STEMI and NSTEMI separately.<sup>5–10</sup>

In most cases, AMI is caused by a ruptured atherosclerotic plaque or erosion of the coronary artery endothelium, triggered by inflammation and followed by thrombosis. STEMI occurs when the thrombosis completely blocks the coronary artery, while NSTEMI occurs when the thrombus is not fully occlusive.<sup>1,3</sup> STEMI accounts for approximately 25–40% of all AMIs and is associated with higher short-term mortality rates, while NSTEMI is more common in older patients with a higher prevalence of comorbidities.<sup>1,3,4</sup> The management of STEMI involves urgent reperfusion therapy with primary percutaneous coronary intervention (PCI) within a few hours of symptom onset. On the other hand, NSTEMI is managed with antiplatelet, anticoagulant, and statin therapies, and urgent PCI is reserved for high-risk cases.<sup>3,4,8,10</sup>

Administrative claims data refers to electronic health data collected and maintained by health insurers and public programs for administrative purposes. It typically includes patient demographics, diagnoses, procedures, prescriptions, and other longitudinal healthcare-related information. For example, Taiwan's National Health Insurance (NHI) claims database is one of the world's largest comprehensive administrative claims databases.<sup>11–13</sup> Administrative claims data has been widely used in cardiovascular research. However, there are limitations to its use, such as potential inaccuracies in diagnosis coding, which can impact the reliability of research findings. Therefore, it is essential to validate claims data by comparing it with primary medical data and definitive diagnoses to ensure the quality and reliability of the data used in research.

In 2014, Cheng et al conducted a study to evaluate the accuracy of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnostic code (410) for identifying patients with AMI in Taiwan's NHI claims data.<sup>14</sup> They found a sensitivity of 0.88 and a positive predictive value (PPV) of 0.92 for identifying AMI patients with a first diagnostic code of 410. However, the accuracy of individual ICD diagnostic codes for STEMI and NSTEMI in the claims data has not been verified. Additionally, since 2016, Taiwan's NHI claims database has switched to using the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) diagnostic codes, which provide more specific classification for STEMI and NSTEMI. Given the distinct epidemiological profiles, pathophysiology, severity, and management approaches of STEMI and NSTEMI, it is crucial to separately examine these two types of AMI when utilizing nationwide real-world claims data. However, there is a lack of research validating the accuracy of ICD-10-CM codes in identifying STEMI and NSTEMI patients in Taiwan's NHI claims database. To address this gap, the current study aims to develop and validate case definition algorithms separately for STEMI and NSTEMI based on ICD-10-CM codes in Taiwan's NHI claims database, using data from two hospitals in Taiwan.

## Materials and Methods

### Study Settings

This retrospective study was conducted at two hospitals in Taiwan that have a contractual agreement with Taiwan's NHI program. The NHI program covers almost all hospitals in Taiwan and requires NHI-contracted hospitals to submit claims in a standardized format to the NHI Administration. The two participating hospitals were among the 46 accredited advanced emergency-responsibility hospitals in the country that offer 24/7 emergency cardiac catheterization services.<sup>15,16</sup>

To develop case definition algorithms, we obtained a dataset from the Ditmanson Medical Foundation Chia-Yi Christian Hospital, a private tertiary teaching hospital with 1000 beds and an annual volume of approximately 500 admissions for AMI. For external validation of these algorithms, we obtained a dataset from the Buddhist Dalin Tzu Chi General Hospital, a private tertiary hospital with 800 beds and approximately 200 AMI admissions per year. We selected this hospital for external validation because it is part of a large healthcare organization that includes eight hospitals across both urban and rural areas in Taiwan that use the same electronic medical record (EMR) system and health information system.

The study protocol was reviewed and approved independently by the institutional review boards of both hospitals, with a waiver of informed consent (approval numbers: Ditmanson Medical Foundation Chia-Yi Christian Hospital IRB2022088 and Buddhist Dalin Tzu Chi General Hospital B11201017).

## Taiwan's National Health Insurance Claims Database

Taiwan's NHI program was launched in 1995 as a compulsory single-payer healthcare system to provide affordable healthcare to all residents. It includes a wide range of services, such as inpatient care, outpatient care, laboratory testing, prescriptions, dental services, home care, and preventative services. The NHI claims database contains the longitudinal claims of nearly 99.8% of Taiwan's population, making it a valuable resource for research on disease prevalence, healthcare policies, and long-term prognosis.<sup>11,17</sup> The NHI claims data includes outpatient claims, inpatient claims, prescriptions dispensed at pharmacies, and registries for beneficiaries, medical facilities, and board-certified specialists. These datasets can be linked with encrypted personal identification numbers to provide patient-level information on demographic characteristics.<sup>11</sup> The NHI claims database released for research allows a maximum of three diagnoses on outpatient claims and five diagnoses on inpatient claims. Diagnoses are coded using the ICD-9-CM until 2015 and the ICD-10-CM from 2016 onwards.

## Data Sources and Participants

The case definition algorithms were developed using data obtained from the Ditmanson Research Database, a deidentified research database containing both administrative claims data and EMRs. This database includes clinical information for approximately 1.4 million patients who received healthcare at the Ditmanson Medical Foundation Chia-Yi Christian Hospital between January 2006 and December 2022, with 0.6 million inpatient and 21.5 million outpatient records. To develop case definition algorithms for STEMI and NSTEMI based on ICD-10-CM diagnostic codes, we extracted claims data and medical records for all inpatients treated between January 2016 and December 2021. For external validation of the algorithms, we used data extracted from the EMRs and claims data of all inpatients treated at the Buddhist Dalin Tzu Chi General Hospital during the same period.

## Case Definition Algorithms

Table 1 presents the algorithms used to identify STEMI, NSTEMI and AMI from the NHI claims data. The algorithms relied on the presence of ICD diagnostic codes ([Supplementary Table 1](#)). Variations to the primary algorithm included: 1) the position of the ICD codes, such as the primary diagnosis field, the first three diagnosis fields, or any diagnosis field, 2) the presence of billing codes for cardiac catheterization or coronary angiography (18020B, 18021B or 18022B), and 3) the presence of a billing code for percutaneous coronary dilatation (33076B, 33077B or 33078B).

## Ascertainment of ST-Elevation and Non-ST-Elevation Myocardial Infarction

To identify all patients with STEMI and NSTEMI, we used three screening methods: keyword search, ICD code search, and procedure code search, as shown in [Figure 1](#). For the keyword search, we screened the diagnoses recorded in the discharge summary of all inpatients during the study period for keywords related to AMI ([Supplementary Table 2](#)) to identify potential AMI cases. The ICD and procedure code search utilized ICD diagnostic codes related to AMI ([Supplementary Table 1](#)) and procedure codes associated with coronary angiography and interventions ([Supplementary Table 3](#)) to identify potential AMI cases from all inpatient claims during the study period.

We manually reviewed the medical records, serial troponin levels, serial ECG reports, and coronary angiography reports of the identified potential AMI cases to determine whether each patient had a true STEMI or NSTEMI, based on the newest universal definition of myocardial infarction.<sup>2,10</sup> In brief, to diagnose AMI, several criteria were considered. First, the patient's cardiac troponin value needed to exceed the 99th percentile upper reference limit, and this value had to change by over 20% in serial examination. Second, the patient's medical records showed symptoms and/or signs related to clinical myocardial ischemia. Third, the ECG exhibited new ischemic or ST-segment changes. Fourth, the patient's coronary angiography indicated the presence of a coronary thrombus or occlusion of the coronary artery. The differential diagnosis between STEMI and NSTEMI was primarily based on the ECG manifestations.<sup>2</sup> The diagnostic criteria were

**Table I** Case Definition Algorithms for STEMI, NSTEMI, and AMI

Diagnosis	Algorithm	ICD Codes	Primary Diagnosis Field	First Three Diagnosis Fields	Any Diagnosis Field	AND	AND
						Cardiac Catheterization / Coronary Angiography Code (18020B, 18021B, 18022B)	Percutaneous Coronary Dilatation Code (33076B, 33077B, 33078B)
<b>STEMI</b>	<b>A</b> <b>B</b> <b>C</b> <b>D</b> <b>E</b> <b>F</b> <b>G</b> <b>H</b> <b>I</b>	STEMI codes	X	X	X		
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
<b>NSTEMI</b>	<b>A</b> <b>B</b> <b>C</b> <b>D</b> <b>E</b> <b>F</b> <b>G</b> <b>H</b> <b>I</b>	NSTEMI codes	X	X	X		
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
<b>AMI</b>	<b>A</b> <b>B</b> <b>C</b> <b>D</b> <b>E</b> <b>F</b> <b>G</b> <b>H</b> <b>I</b>	STEMI and NSTEMI codes	X	X	X		
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	
			X	X		X	

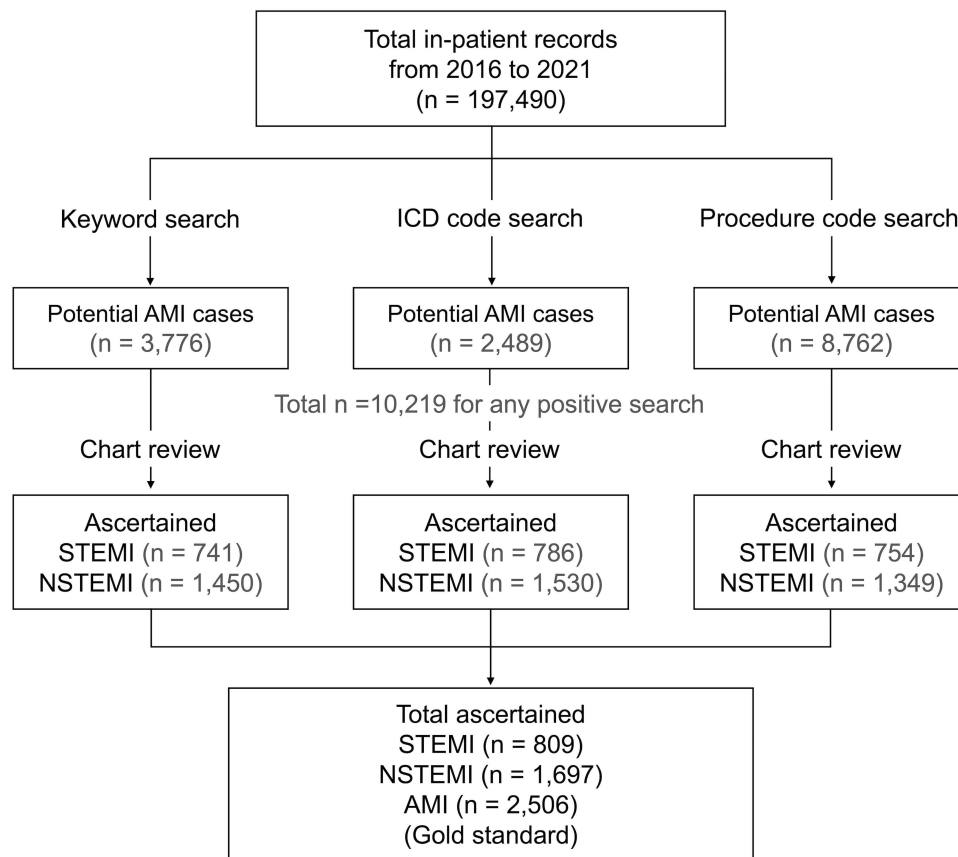
**Abbreviations:** AMI, acute myocardial infarction; ICD, International Classification of Diseases diagnostic codes; STEMI, ST-elevation myocardial infarction; NSTEMI, non-ST-elevation myocardial infarction.

shown in [Supplementary Table 4](#). Two investigators (MJT and JHL) conducted the review, and any discrepancies were resolved by consensus or by a third investigator (JFL). The consensus diagnosis was considered the gold standard for all analyses ([Figure 1](#)).

## Analysis and External Validation

We evaluated the performance of each algorithm in identifying true STEMI and NSTEMI cases by calculating sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) based on the confusion matrix ([Figure 2](#)). We also calculated their 95% confidence intervals (CIs) using the Clopper-Pearson exact method. The primary evaluation metric was the kappa coefficient, which was used to measure the agreement between algorithm-detected AMI cases and true AMI patients. We interpreted the degree of agreement as follows: “slight” (0.00–0.20), “fair” (0.21–0.40), “moderate” (0.41–0.60), “substantial” (0.61–0.80), and “perfect” (0.81–1.00).<sup>18</sup>

We externally validated the algorithms with the highest kappa coefficients on an independent dataset from the Buddhist Dalin Tzu Chi General Hospital. A similar process was undertaken to identify potential AMI cases, and one investigator (TYT) determined if each patient was a true STEMI or NSTEMI using the same criteria. Additionally, we



**Figure 1** Flow chart outlining the verification process of true AMI cases.

**Abbreviations:** AMI, acute myocardial infarction; ICD, International Classification of Diseases diagnostic codes; STEMI, ST-elevation myocardial infarction; NSTEMI, non-ST-elevation myocardial infarction.

performed subgroup analysis by stratifying the data based on the year of admission and sex to examine any variations in accuracy among different subgroups. Subgroup differences in sensitivity and PPV were assessed using a likelihood ratio test. All data analyses were conducted using Stata 17.0 (StataCorp, College Station, Texas), and we considered statistical significance to be two-tailed  $p < 0.05$ .

## Results

During the study period, 197,490 inpatient visits were retrieved from the hospital inpatient claims. The keyword search identified 3776 potential cases of AMI, of which 741 were adjudicated as true STEMI and 1450 as true NSTEMI through manual chart review (Figure 1). The ICD code search identified 2489 potential cases of AMI, of which 786 were adjudicated as true STEMI and 1530 as true NSTEMI through chart review. The procedure code search identified 8762 potential cases of AMI, of which 754 were adjudicated as true STEMI and 1349 as true NSTEMI. In total, 809 cases of STEMI and 1697 cases of NSTEMI were ascertained during the study period.

The validity results of the best and all algorithms for identifying STEMI, NSTEMI and AMI are presented in Table 2 and Supplementary Table 5, respectively. The best algorithm for identifying STEMI cases was algorithm B. It defined STEMI as the presence of any STEMI ICD code (ICD-10-CM codes: I21.0, I21.1, I21.2, I21.3, I21.9) in the first three diagnosis fields. Its sensitivity, specificity, PPV, NPV, and kappa were 93.6%, 100%, 89.4%, 100%, and 0.914, respectively. For identifying NSTEMI cases, algorithm C was the best. It defined NSTEMI as the presence of the NSTEMI ICD code (ICD-10-CM codes: I21.4) in any diagnosis fields. Its sensitivity, specificity, PPV, NPV, and kappa were 82.6%, 100%, 96.5%, 99.8%, and 0.889, respectively. The best algorithm for identifying AMI cases (including STEMI and NSTEMI) was algorithm C. It defined AMI

		STEMI or NSTEMI ascertained by experts (Gold standard)	
		Positive	Negative
STEMI or NSTEMI identified by algorithms	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{PPV} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$$

$$\text{NPV} = \text{TN} / (\text{TN} + \text{FN})$$

**Figure 2** Sample validation matrix analyzing algorithm performance against the reference standard.

**Abbreviations:** PPV, positive predictive value; NPV, negative predictive value; STEMI, ST-elevation myocardial infarction; NSTEMI, non-ST-elevation myocardial infarction.

as the presence of either STEMI or NSTEMI ICD codes in any diagnosis fields. Its sensitivity, specificity, PPV, NPV, and kappa were 89.4%, 99.9%, 95.6%, 99.9%, and 0.923, respectively.

Incorporating billing codes for coronary angiography and interventions in the algorithms led to a minor increase in PPV and a minor decrease in sensitivity for STEMI case detection ([Supplementary Table 5](#), STEMI algorithms E to I). However, this resulted in a marked decrease in sensitivity and kappa for detecting NSTEMI and AMI cases, despite a slight improvement in PPV ([Supplementary Table 5](#), NSTEMI and AMI algorithms E to I).

**Table 3** summarizes the reasons for false-positive and false-negative identifications of STEMI, NSTEMI, and AMI by their respective best algorithms. In STEMI algorithm B, the primary reasons for false negatives were STEMI codes being listed in the 4th or 5th diagnosis field and STEMI cases being miscoded as atherosclerotic heart disease (I25.1) or underlying diseases. False positives for algorithm B in detecting STEMI cases were mainly caused by cases of NSTEMI or unstable angina being miscoded as STEMI. Regarding NSTEMI algorithm C, the main reasons for false negatives

**Table 2** Performance of the Best Algorithms in Identifying STEMI, NSTEMI, and AMI Among All in-Patients (n = 197,490) in the Development Dataset

Diagnosis	Best Performing Algorithm		Ascertained Diagnosis		Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)	Kappa (95% CI)
			Positive	Negative					
STEMI	Alg B	Positive	757	90	93.6	100.0	89.4	100.0	0.914
		Negative	52	196,591	(91.7–95.2)	(99.9–100.0)	(87.1–91.4)	(100.0–100.0)	(0.900–0.928)
NSTEMI	Alg C	Positive	1402	51	82.6	100.0	96.5	99.8	0.889
		Negative	295	195,742	(80.7–84.4)	(100.0–100.0)	(95.4–97.4)	(99.8–99.9)	(0.878–0.901)
AMI	Alg C	Positive	2241	104	89.4	99.9	95.6	99.9	0.923
		Negative	265	194,880	(88.2–90.6)	(99.9–100.0)	(94.7–96.4)	(99.8–99.9)	(0.915–0.931)

**Abbreviations:** Alg, algorithm; AMI, acute myocardial infarction; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; STEMI, ST-elevation myocardial infarction; NSTEMI, non-ST-elevation myocardial infarction.



**Table 3** Reasons for False-Negative and False-Positive Identification of STEMI, NSTEMI and AMI by the Best Algorithms

Diagnosis	Reason	n
STEMI	<b>False negatives for algorithm B</b>	<b>52</b>
	STEMI codes listed in the 4th or 5th diagnosis fields	17
	Miscoded as atherosclerotic heart disease of native coronary artery (I25.IX)	16
	Coded as comorbidities or underlying diseases instead of STEMI	14
	Miscoded as acute ischemic heart diseases, unspecified (I24.9)	5
	<b>False positives for algorithm B</b>	<b>90</b>
	NSTEMI miscoded as STEMI	58
	Unstable angina miscoded as STEMI	22
	Suspected coronary artery disease without coronary angiography confirmation	9
	Coronary artery disease with chronic stable angina miscoded as STEMI	1
NSTEMI	<b>False negatives for algorithm C</b>	<b>295</b>
	Coded as comorbidities or underlying diseases instead of NSTEMI	153
	Miscoded as atherosclerotic heart disease of native coronary artery (I25.IX)	70
	Miscoded as STEMI	56
	Miscoded as acute ischemic heart diseases, unspecified (I24.9)	15
	Miscoded as subsequent NSTEMI (I22.2)	1
	<b>False positives for algorithm C</b>	<b>51</b>
	Unstable angina miscoded as NSTEMI	29
	Suspected coronary artery disease without coronary angiography confirmation	18
	STEMI miscoded as NSTEMI	2
Coronary artery disease with chronic stable angina miscoded as NSTEMI	1	
Negative coronary angiography finding miscoded as NSTEMI	1	
AMI	<b>False negatives for algorithm C</b>	<b>265</b>
	Coded as comorbidities or underlying diseases instead of STEMI or NSTEMI	160
	Miscoded as atherosclerotic heart disease of native coronary artery (I25.IX)	84
	Miscoded as acute ischemic heart diseases, unspecified (I24.9)	20
	Miscoded as subsequent NSTEMI (I22.2)	1
	<b>False positives for algorithm C</b>	<b>104</b>
	Unstable angina miscoded as STEMI or NSTEMI	62
	Suspected coronary artery disease without coronary angiography confirmation	35
	Negative coronary angiography finding miscoded as STEMI or NSTEMI	5
	Coronary artery disease with chronic stable angina miscoded as NSTEMI	2

**Abbreviations:** AMI, acute myocardial infarction; STEMI, ST-elevation myocardial infarction; NSTEMI, non-ST-elevation myocardial infarction.

were being miscoded as comorbidity, underlying diseases, atherosclerotic heart disease, or STEMI. False positives in NSTEMI algorithm C were caused by unstable angina cases or suspicious cases without coronary angiography confirmation being miscoded as NSTEMI. Similar reasons for false negative and false positive results in using algorithm C in detecting AMI cases were also found, as listed in Table 3.

The most effective algorithms for identifying STEMI, NSTEMI, and AMI were validated using an external dataset from a different hospital, which included 152,793 inpatient visits. Table 4 shows the sensitivity, specificity, PPV, NPV, and kappa values of these algorithms. The results demonstrated that these algorithms effectively identified STEMI, NSTEMI, and AMI among the external cohort.

To assess whether the coding accuracy of the best individual algorithms was affected by the year of admission or sex, we conducted subgroup analyses stratified by the year of admission (Figure 3) and sex (Table 5). The performance of the STEMI algorithm B remained consistent throughout the years, with no significant changes in sensitivity ( $p = 0.312$ ) and PPV ( $p = 0.317$ ) from 2016 to 2021 (Figure 3). However, significant differences in sensitivity were observed for the

**Table 4** Performance of the Best Algorithms in Identifying STEMI, NSTEMI, and AMI Among All in-Patients (n = 152,793) in the Validation Dataset

Diagnosis	Best Performing Algorithm		Ascertained Diagnosis		Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)	Kappa (95% CI)
			Positive	Negative					
STEMI	Alg B	Positive	296	53	95.2	100.0	84.8	100.0	0.897
		Negative	15	152,429	(92.8–97.6)	(100.0–100.0)	(81.1–88.6)	(100.0–100.0)	(0.872–0.921)
NSTEMI	Alg C	Positive	648	84	96.3	99.9	88.5	100.0	0.922
		Negative	25	152,036	(94.9–97.7)	(99.9–100.0)	(86.2–90.8)	(100.0–100.0)	(0.907–0.937)
AMI	Alg C	Positive	964	131	98	99.9	88.0	100.0	0.927
		Negative	20	151,678	(97.1–98.9)	(99.9–99.9)	(86.1–90.0)	(100.0–100.0)	(0.915–0.939)

**Abbreviations:** Alg, algorithm; AMI, acute myocardial infarction; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; STEMI, ST-elevation myocardial infarction; NSTEMI, non-ST-elevation myocardial infarction.

NSTEMI algorithm C ( $p = 0.011$ ) and AMI algorithm C ( $p = 0.004$ ) among the different year subgroups (Figure 3). Regarding sex subgroups, the algorithms generally performed better in male patients compared to female patients. Significant differences were found in the sensitivity of the STEMI algorithm B ( $p = 0.014$ ) and AMI algorithm C ( $p = 0.001$ ), as well as the PPV of the AMI algorithm C ( $p = 0.038$ ) (Table 5).

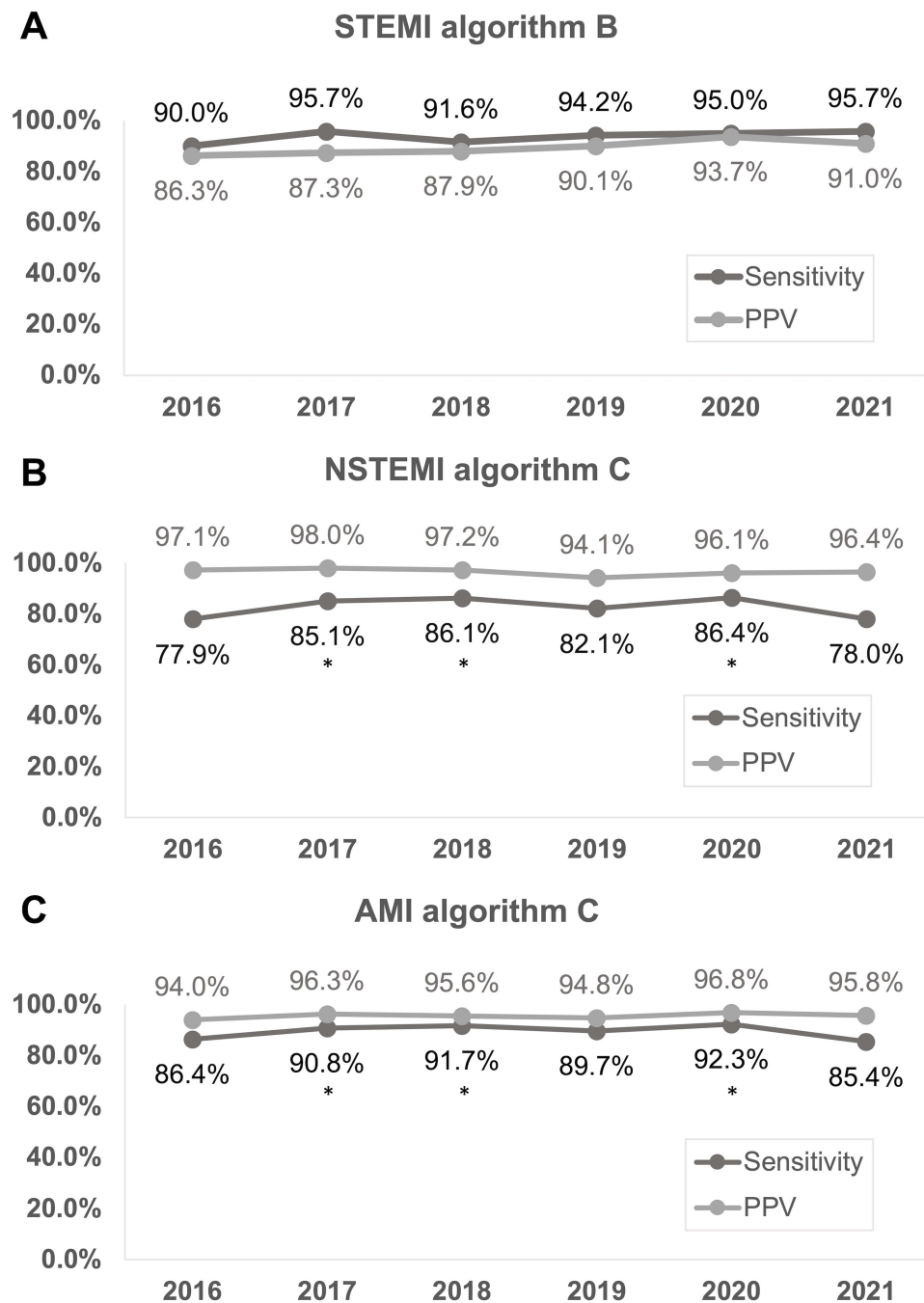
## Discussion

In this study, we developed and evaluated case definitions for STEMI, NSTEMI, and AMI separately using ICD-10-CM codes and billing codes for coronary angiography and interventions in Taiwan's NHI claims database. The algorithm that defined STEMI using STEMI ICD codes in the first three diagnosis fields demonstrated the best performance. For identifying NSTEMI, the algorithm utilizing the NSTEMI ICD code listed in any diagnosis field exhibited superior performance. The algorithm that included either STEMI or NSTEMI ICD codes listed in any diagnosis field demonstrated excellent performance in defining AMI. All three algorithms exhibited satisfactory performance during external validation.

Over the past two decades, there has been a significant increase in research utilizing administrative databases to investigate cardiovascular diseases. The number of publications in this field has substantially grown, from only two publications in 1996 to 268 publications in 2018.<sup>13</sup> Furthermore, these publications have received considerable attention, as evidenced by their citation rates.<sup>13</sup> However, before utilizing administrative databases for epidemiological and population health research, it is crucial to develop and validate case definitions.

Several studies have examined the validity of ICD codes for diagnosing AMI in various health administrative databases, including regional and national registries. For example, Kiyota et al (2004) validated the ICD-9-CM discharge diagnostic codes (410.X0 and 410.X1) for AMI in the US Medicare system. They reported a PPV of 94.1% when considering hospitalization for at least 3 days and diagnostic codes in the principal or secondary diagnostic field.<sup>19</sup> Similarly, Kimm et al assessed the validity of ICD-10-CM codes (I21.x) in the Korean National Medical Health Insurance claims data. They found PPVs ranging from 71.4% to 73.1% based on different criteria.<sup>20</sup> Nakai et al conducted a validation study using the Japanese Registry Of All cardiac and vascular Disease database. They demonstrated satisfactory accuracy in identifying AMI using the ICD-10-CM code (I21.x), with a sensitivity of 78.9% and a PPV of 78.8%. The PPV increased to 84.9% when emergency hospitalization was considered.<sup>21</sup> Based on these findings, hospitalization is an important criterion for improving the accuracy of AMI diagnosis using ICD codes. Therefore, our study focused on hospitalized patients and utilized inpatient claim data to define AMI. It is essential to note that our findings for defining AMI will be limited to studies that use inpatient claim data from Taiwan's NHI database.





**Figure 3** The sensitivity and PPV of the best algorithms for identifying STEMI (A), NSTEMI (B) and AMI (C) from 2016 to 2021.

**Note:** \*Significant difference ( $P < 0.05$ ) compared to the reference year of 2016.

**Abbreviations:** PPV, positive predictive value; STEMI, ST-elevation myocardial infarction; NSTEMI, non-ST-elevation myocardial infarction.

A recent analysis, using bibliometric and text mining techniques on PubMed, examined the prevalent research topics based on Taiwan's NHI claims database.<sup>12</sup> The findings revealed that, unlike stroke, AMI is not a primary focus of research utilizing Taiwan's NHI claims database. This disparity suggests that the limited emphasis on AMI research may be attributed to inadequate definition and validation of STEMI and NSTEMI within the database. A previous study assessed the accuracy of using ICD-9-CM diagnostic codes to identify AMI patients in Taiwan's NHI claims data, demonstrating satisfactory validity (sensitivity: 0.88, PPV: 0.92).<sup>14</sup> However, it is noteworthy that Taiwan's NHI database has transitioned to the ICD-10-CM coding system since 2016, and there is currently a lack of validation research in this

**Table 5** Performance of the Best Algorithms Among Different Subgroups Stratified by Sex

Subgroup	Best Performing Algorithm	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)	Kappa (95% CI)
Male patients (n = 101,300)	STEMI Alg B	94.8* (92.8–96.4)	99.9 (99.9–100.0)	90.1 (87.6–92.3)	100.0 (100.0–100.0)	0.924 (0.908–0.939)
	NSTEMI Alg C	83.5 (81.2–85.6)	100.0 (100.0–100.0)	97.0 (95.7–98.0)	99.8 (99.8–99.8)	0.896 (0.882–0.910)
	AMI Alg C	90.8* (89.3–92.1)	99.9 (99.9–100.0)	96.2* (95.1–97.0)	99.8 (99.8–99.9)	0.933 (0.924–0.942)
Female patients (n = 96,190)	STEMI Alg B	89.6 (84.4–93.5)	100.0 (100.0–100.0)	86.9 (81.4–91.2)	100.0 (100.0–100.0)	0.882 (0.848–0.916)
	NSTEMI Alg C	80.9 (77.5–84.1)	100.0 (100.0–100.0)	95.5 (93.3–97.2)	99.9 (99.9–99.9)	0.875 (0.854–0.897)
	AMI Alg C	86.3 (83.7–88.7)	100 (99.9–100.0)	94.2 (92.2–95.8)	99.9 (99.9–99.9)	0.9 (0.884–0.916)

**Note:** \*Significant difference ( $P < 0.05$ ) compared to the corresponding female subgroup.

**Abbreviations:** Alg, algorithm; AMI, acute myocardial infarction; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; STEMI, ST-elevation myocardial infarction; NSTEMI, non-ST-elevation myocardial infarction.

context. Additionally, there is a dearth of studies examining the validity of ICD codes for the individual diagnosis of STEMI and NSTEMI. Therefore, we believe that this study will provide a critical and meaningful reference for future AMI research using Taiwan's NHI claims database.

Previous studies have shown that transitioning from one coding system to another can lead to inconsistent estimation of disease prevalence when using diagnostic codes.<sup>22,23</sup> In 2016, the NHI in Taiwan switched from the ICD-9-CM to ICD-10-CM coding system. Our previous study indicated that this transition had an impact on the validity of identifying out-of-hospital cardiac arrest cases using ICD-10-CM-based algorithms, with an observable decrease in validity in the first year after the transition.<sup>24</sup> This might be due to unfamiliarity with the new coding system and potential coding errors. To examine whether a similar issue existed in the development of our algorithms, we performed a subgroup analysis to assess the accuracy of the algorithm in different years. Figure 3 shows that the STEMI algorithm demonstrated consistent performance across different years. However, we observed a significantly lower sensitivity in the year 2016 compared to the years 2017, 2018, and 2020 for both our NSTEMI and AMI algorithms. This suggests that the transition to the ICD-10-CM coding system may have affected the validity of our NSTEMI and AMI algorithms. Nevertheless, the performance of the two algorithms is still acceptable. However, researchers should remain cautious regarding the decrease in validity observed in the first year following the coding system transition.

Prior research has highlighted sex disparities in the characteristics, pathophysiology, presentation, management, and outcomes of STEMI and NSTEMI.<sup>3,25</sup> To evaluate the potential impact of sex on the accuracy of our developed algorithms, we conducted another subgroup analysis (Table 5). The results revealed differences in sensitivity for the STEMI and AMI algorithms, as well as the PPV for the AMI algorithm, between males and females. Generally, the algorithms performed better in male patients compared to female patients. However, it is worth mentioning that, overall, our developed algorithms demonstrated satisfactory performance for both sexes.

In our study, we noted false-positive identification of STEMI or NSTEMI cases, primarily due to miscoding as the opposite type of AMI, unstable angina, or suspected cases without confirmation through coronary angiography. To ensure the accuracy of our study, we specifically defined AMI cases based on evidence from coronary angiography, including the presence of a thrombus or occlusion. Therefore, cases without confirmation through coronary angiography were not considered true AMI cases. Although this approach may have slightly reduced the PPV, it ensures the reliability of our results. On the other hand, false-negative identification of different types of AMI primarily occurred when cases were miscoded as comorbidities, underlying diseases, or atherosclerotic heart disease instead of the correct type of AMI. This

issue was more prominent in identifying NSTEMI cases, leading to a reduction in sensitivity for detecting NSTEMI cases. However, it is important to note that NSTEMI cases differ from STEMI cases. A significant proportion of NSTEMI patients have other comorbidities or NSTEMI occurs as a secondary condition to other primary diseases.<sup>1,3,4</sup> This complexity in NSTEMI diagnosis might account for the observed reduction in sensitivity, which is reasonable given the multifaceted nature of NSTEMI.

This study has several limitations. Firstly, the algorithms were developed using data from only one hospital, as in previous studies.<sup>14,19,26</sup> Therefore, it is unclear whether the findings can be generalized to the entire population covered by Taiwan's NHI. However, all hospitals in Taiwan are required to be accredited and follow the same data processing systems for reimbursement. At a minimum, the developed algorithms performed well on an external dataset from another accredited advanced emergency-responsibility hospital. This suggests that our findings are likely to be applicable to the 46 accredited advanced emergency-responsibility hospitals in Taiwan. Secondly, the algorithms for identifying STEMI, NSTEMI, and AMI can only be applied to studies using Taiwan's NHI database. Future studies may be required to explore the applicability of these algorithms in databases from other healthcare systems with similar coding practices. To the best of our knowledge, there is a lack of research validating the accuracy of ICD-10-CM codes in identifying STEMI and NSTEMI patients separately in other national claims databases. Therefore, the methods used in this study could serve as a reference for developing algorithms in other healthcare systems for future research. Thirdly, this study only validated the individual ICD-10-CM diagnostic codes for STEMI and NSTEMI, and the findings may not apply to the ICD-9-CM code era in Taiwan's NHI database. Fourthly, the algorithms developed in this study only included the procedure codes, such as coronary angiography and intervention. Further studies may be needed to investigate whether adding medication codes, length of hospital stay, or the condition of post-hospital discharge follow-up could improve the diagnostic performance. Lastly, the universal definition of myocardial infarction may change with advancements in medical testing. Therefore, our results may not be applicable if there are changes in the definition of AMI in the future.

## Conclusion

In Taiwan, a national acute coronary syndrome registry existed in the past, but it only covered limited years (2008–2010 and 2012–2015) and did not include hospitals nationwide.<sup>27,28</sup> Therefore, if future studies aim to investigate the impact of new drugs, interventions, or public health events such as the COVID-19 pandemic on different types of AMI, the NHI database will be a valuable resource. However, before conducting epidemiological studies using claims databases, it is essential to validate the case definitions of the disease of interest. Currently, Taiwan's NHI claims database lacks validated case definitions for STEMI and NSTEMI as separate entities. In this study, we developed and evaluated multiple case definition algorithms for STEMI and NSTEMI using ICD-10-CM diagnostic codes and billing codes for NHI reimbursement. The results of this study will provide valuable guidance for future researchers, allowing them to use appropriate case definitions for STEMI and NSTEMI that align with their specific research objectives.

## Ethics Approval and Informed Consent

This study was conducted in accordance with the Declaration of Helsinki and received independent approval from the Institutional Review Board of Ditmanson Medical Foundation Chia-Yi Christian Hospital (IRB 2022088) and the Institutional Review Board of Buddhist Dalin Tzu Chi General Hospital (B11201017). Both institutional review boards approved the exemption of patient informed consent for this retrospective study. The study data were treated with confidentiality to protect the privacy of the participants.

## Acknowledgments

Tou-Yuan Tsai and Jen-Feng Lin are co-first authors for this study. The authors thank the help from the Clinical Data Center, Ditmanson Medical Foundation Chia-Yi Christian Hospital for providing administrative and technical support. This study is based in part on data from the Ditmanson Research Database (DRD) provided by Ditmanson Medical Foundation Chia-Yi Christian Hospital. The interpretation and conclusions contained herein do not represent the position of Ditmanson Medical Foundation Chia-Yi Christian Hospital.

## Funding

This research was funded by the Ditmanson Medical Foundation Chia-Yi Christian Hospital Research Program (grant number R111-19). The funders of the research had no role in the design and conduct of the study, interpretation of the data, or decision to submit for publication.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Reed GW, Rossi JE, Cannon CP. Acute myocardial infarction. *Lancet*. 2017;389(10065):197–210. doi:10.1016/S0140-6736(16)30677-8
2. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). *J Am Coll Cardiol*. 2018;72(18):2231–2264. doi:10.1016/j.jacc.2018.08.1038
3. Bhatt DL, Lopes RD, Harrington RA. Diagnosis and treatment of acute coronary syndromes: a review. *JAMA*. 2022;327(7):662–675. doi:10.1001/jama.2022.0358
4. Mitsis A, Gragnano F. Myocardial Infarction with and without ST-segment elevation: a contemporary reappraisal of similarities and differences. *Curr Cardiol Rev*. 2021;17(4):e230421189013. doi:10.2174/1573403X16999201210195702
5. O’Gara PT, Kushner FG, Ascheim DD, et al. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines. *Circulation*. 2013;127(4):e362–425. doi:10.1161/CIR.0b013e3182742c84
6. Levine GN, Bates ER, Blankenship JC, et al. 2015 ACC/AHA/SCAI focused update on primary percutaneous coronary intervention for patients with ST-elevation myocardial infarction: an update of the 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention and the 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction. *J Am Coll Cardiol*. 2016;67(10):1235–1250. doi:10.1016/j.jacc.2015.10.005
7. Roffi M, Patrono C, Collet JP, et al. 2015 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: task force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J*. 2016;37(3):267–315. doi:10.1093/eurheartj/ehv320
8. Ibanez B, James S, Agewall S, et al. 2017 ESC guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: the task force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J*. 2018;39(2):119–177. doi:10.1093/eurheartj/ehx393
9. Li YH, Lee CH, Huang WC, et al. 2020 focused update of the 2012 guidelines of the Taiwan Society of cardiology for the management of ST-segment elevation myocardial infarction. *Acta Cardiol Sin*. 2020;36(4):285–307. doi:10.6515/ACS.202007\_36(4).20200619A
10. Collet JP, Thiele H, Barbato E, et al. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. *Eur Heart J*. 2021;42(14):1289–1367. doi:10.1093/eurheartj/ehaa575
11. Hsieh CY, Su CC, Shao SC, et al. Taiwan’s national health insurance research database: past and future. *Clin Epidemiol*. 2019;11:349–358. doi:10.2147/CLEP.S196293
12. Sung SF, Hsieh CY, Hu YH. Two decades of research using Taiwan’s national health insurance claims data: bibliometric and text mining analysis on PubMed. *J Med Internet Res*. 2020;22(6):e18457. doi:10.2196/18457
13. AlHajji M, Alqahtani F, Alkhouli M. Contemporary trends in the utilization of administrative databases in cardiovascular research. *Mayo Clin Proc*. 2019;94(6):1120–1121. doi:10.1016/j.mayocp.2019.03.022
14. Cheng CL, Lee CH, Chen PS, Li YH, Lin SJ, Yang YH. Validation of acute myocardial infarction cases in the national health insurance research database in Taiwan. *J Epidemiol*. 2014;24(6):500–507. doi:10.2188/jea.JE20140076
15. Joint commission of Taiwan-accreditation and certification-accreditation, inspection, certification-hospital accreditation and certification. Available from: <https://www.jct.org.tw/cp-1339-8352-44faf-2.html>. Accessed December 20, 2022.
16. Lin CY, Lee YC. Effectiveness of hospital emergency department regionalization and categorization policy on appropriate patient emergency care use: a nationwide observational study in Taiwan. *BMC Health Serv Res*. 2021;21(1):21. doi:10.1186/s12913-020-06006-7
17. National Health Insurance Administration. NHI profile. Universal Health Coverage in Taiwan. Available from: [https://www.nhi.gov.tw/English/Content\\_List.aspx?n=4D7051840BF42F52&topn=ED4A30E51A609E49](https://www.nhi.gov.tw/English/Content_List.aspx?n=4D7051840BF42F52&topn=ED4A30E51A609E49). Accessed February 24, 2022.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. doi:10.2307/2529310
19. Kiyota Y, Schneeweiss S, Glynn RJ, Cannuscio CC, Avorn J, Solomon DH. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. *Am Heart J*. 2004;148(1):99–104. doi:10.1016/j.ahj.2004.02.013
20. Kimm H, Yun JE, Lee SH, Jang Y, Jee SH. Validity of the diagnosis of acute myocardial infarction in Korean national medical health insurance claims data: the Korean heart study (1). *Korean Circ J*. 2012;42(1):10–15. doi:10.4070/kcj.2012.42.1.10
21. Nakai M, Iwanaga Y, Sumita Y, et al. Validation of acute myocardial infarction and heart failure diagnoses in hospitalized patients with the nationwide claim-based JROAD-DPC database. *Circ Rep*. 2021;3(3):131–136. doi:10.1253/circrep.CR-21-0004
22. Hsu MC, Wang CC, Huang LY, Lin CY, Lin FJ, Toh S. Effect of ICD-9-CM to ICD-10-CM coding system transition on identification of common conditions: an interrupted time series analysis. *Pharmacoepidemiol Drug Saf*. 2021;30(12):1653–1674. doi:10.1002/pds.5330
23. Metcalfe A, Sheikh M, Hetherington E. Impact of the ICD-9-CM to ICD-10-CM transition on the incidence of severe maternal morbidity among delivery hospitalizations in the United States. *Am J Obstet Gynecol*. 2021;225(4):422 e421–422 e411. doi:10.1016/j.ajog.2021.03.036
24. Tsai MJ, Tsai CH, Pan RC, Hsu CF, Sung SF. Validation of ICD-9-CM and ICD-10-CM diagnostic codes for identifying patients with out-of-hospital cardiac arrest in a national health insurance claims database. *Clin Epidemiol*. 2022;14:721–730. doi:10.2147/CLEP.S366874

25. Ashraf M, Jan MF, Bajwa TK, Carnahan R, Zlochiver V, Allaqaband SQ. Sex disparities in diagnostic evaluation and revascularization in patients with acute myocardial infarction-A 15-year nationwide study. *J Am Heart Assoc.* 2023;12(6):e027716. doi:10.1161/JAHA.122.027716
26. Shima D, Ii Y, Higa S, et al. Validation of novel identification algorithms for major adverse cardiovascular events in a Japanese claims database. *J Clin Hypertens.* 2021;23(3):646–655. doi:10.1111/jch.14151
27. Li YH, Chiu YW, Cheng JJ, et al. Changing practice pattern of acute coronary syndromes in Taiwan from 2008 to 2015. *Acta Cardiol Sin.* 2019;35(1):1–10. doi:10.6515/ACS.201901\_35(1).20180716B
28. Wei CC, Shyu KG, Cheng JJ, Lo HM, Chiu CZ. The relation between the timing of percutaneous coronary intervention and outcomes in patients with acute coronary syndrome with routine invasive strategy - data from Taiwan acute coronary syndrome full spectrum data registry. *Acta Cardiol Sin.* 2016;32(1):39–48. doi:10.6515/acs20150722a

## Clinical Epidemiology

Dovepress

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>