Performance of ICD-10-CM Diagnosis Codes for Identifying Acute Ischemic Stroke in a National Health Insurance Claims Database

Meng-Tsang Hsieh1–3,* Cheng-Yang Hsieh4,5,* Tzu-Tung Tsai1 Yi-Ching Sheng-Feng Sung1,6–8

1Stroke Center and Department of Neurology, E-Da Hospital, Kaohsiung, Taiwan; 2School of Medicine for International Students, College of Medicine, I-Shou University, Kaohsiung, Taiwan; 3Institute of Clinical Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwan; 4Department of Neuroscience, Tainan Sin Lau Hospital, Tainan, Taiwan; 5School of Pharmacy, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan; 6Division of Neurology, Department of Internal Medicine, Dittmanon Medical Foundation Chia-Yi Christian Hospital, Chia-Yi City, Taiwan; 7Department of Information Management and Institute of Healthcare Information Management, National Chung Cheng University, Chiayi County, Taiwan; 8Department of Nursing, Min-Hwei Junior College of Health Care Management, Tainan, Taiwan

*These authors contributed equally to this work

Purpose: The validity of the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) coding for the identification of acute ischemic stroke (AIS) in Taiwan’s National Health Insurance claims database has not been investigated. This study aimed to construct and validate the case definition algorithms for AIS based on ICD-10-CM diagnostic codes.

Patients and Methods: This study identified all hospitalizations with ICD-10-CM code I63* in any position of the discharge diagnoses from the inpatient claims database and all patients with a final diagnosis of AIS from the stroke registry between Jan 2018 and Dec 2019. Hospitalizations in the claims data that could be successfully linked to those in the registry data were regarded as true episodes of AIS. Otherwise, their electronic medical records and images were manually reviewed to ascertain whether they were true episodes of AIS. Using the true episodes of AIS as the reference standard, the positive predictive value (PPV) and sensitivity of various case definition algorithms for AIS were calculated.

Results: A total of 1227 hospitalizations were successfully linked. Among the 155 hospitalizations that could not be linked, 54 were determined to be true episodes of AIS. Using ICD-10-CM code I63* in any position of the discharge diagnoses to identify AIS yielded a PPV and sensitivity of 92.7% and 99.4%, respectively. The PPV increased to 99.8% with >12% decrease in the sensitivity when AIS was restricted to those with I63* as the primary diagnosis. When AIS was defined to be I63* as the primary, first secondary, or second secondary diagnosis, both PPV and sensitivity were greater than 97%.

Conclusion: This study demonstrated the validity of various case definition algorithms for AIS based on ICD-10-CM coding and can provide a reference for future claims-based stroke research.

Keywords: administrative claims data, diagnosis, ICD-10-CM, acute ischemic stroke

Introduction

Stroke remains a major leading cause of death and disability of human beings worldwide.1 Despite a decrease in the age-standardized mortality rates for stroke recently, the global burden of stroke is great and increasing.2 The global lifetime risk of stroke was estimated to be 24.9% for any type of stroke and 18.3% for ischemic stroke in individuals aged 25 years or older.3 In response to the huge and growing burden of stroke worldwide, more studies are required regarding the prevention, early diagnosis, treatment, quality of care, and outcomes of stroke.

Administrative claims data, which are derived from insurance reimbursement records routinely generated during healthcare encounters, have become an...
affordable and widely used data source for researchers to conduct stroke research.\textsuperscript{4,5} For example, stroke is the second most popular research topic among 4473 original articles published between 2000 and 2017 using Taiwan’s National Health Insurance (NHI) claims database.\textsuperscript{6} Because claims databases are designed primarily for reimbursement rather than research purposes, they are not as useful as databases specifically compiled for conducting clinical studies unless their data validity can be verified.\textsuperscript{7} Therefore, confirming the accuracy of case or outcome ascertainment algorithms through validation studies is essential for improving research using administrative claims databases.\textsuperscript{8} To this end, several studies have been performed to establish the validity of ascertainment algorithms for stroke and its risk factors in the NHI claims database based on the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes.\textsuperscript{9-11} In these studies, the sensitivity and positive predictive value (PPV) of the case definition for ischemic stroke were 94.5% to 97.3% and 88.4% to 97.9%, respectively.\textsuperscript{9,10} Consequently, researchers might be more confident about conducting stroke research using the NHI claims database.\textsuperscript{6}

However, as Taiwan switched to the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) coding in 2016,\textsuperscript{12} the previously validated case definition coding in 2016,\textsuperscript{12} the previously validated case definition algorithms are no longer suitable. Therefore, by linking stroke registry data to claims data, this study aimed to construct and validate the case definition algorithms for acute ischemic stroke (AIS) in the NHI claims database based on ICD-10-CM diagnosis codes.

Patients and Methods

Study Settings

This study was conducted using data from the E-Da Hospital, an NHI-contracted would-be medical center in southern Taiwan, which has approximately 1200 beds with an average of 100,000 outpatient and 23,000 inpatient visits per month. The Stroke Center of E-Da Hospital has participated in the Taiwan Stroke Registry (TSR) program\textsuperscript{13} since 2008 and its quality of stroke care has been accredited by the Joint Commission of Taiwan. The study protocol was approved by the Institutional Review Board of E-Da Hospital (EMRP-109-013). The requirement for informed consent was waived because of the retrospective design. The study data were maintained with confidentiality to ensure the privacy of the participants.

Data Sources and Record Linkage

This study used stroke registry data and administrative claims data from the E-Da Hospital. The stroke registry prospectively registered consecutive patients hospitalized for stroke within 10 days of symptom onset in accordance with the design of the TSR.\textsuperscript{13} Patient demographics, medical history, comorbidities, stroke severity as assessed by the National Institutes of Health Stroke Scale, treatments received, hospital course, and final diagnoses were collected. The inpatient claims data reported to the Administration of NHI were retrieved from the hospital information system. Both data sources were linked via patients’ unique personal identifier and the date of admission.

Study Sample

Patients discharged between Jan 2018 and Dec 2019 with a final diagnosis of AIS were identified from the stroke registry. Those with an in-hospital stroke were retained. Patients with missing information on birthday, sex, or personal identifier were excluded. Multiple hospitalizations for the same stroke episode were merged. In contrast, multiple hospitalizations due to different stroke episodes for the same patient were included for analysis. Similarly, all patients discharged during the same period with an ICD-10-CM diagnosis code of I63* (meaning all the diagnosis codes that begin with I63) in any position of the discharge diagnoses were identified from the inpatient claims data. Hospitalization records from either data sources were combined to form the study sample (Figure 1).

Ascertainment of Acute Ischemic Stroke

Figure 1 displays the process of case ascertainment. According to the TSR program, patients presenting within 10 days of symptom onset with acute onset of neurological signs or symptoms persisting longer than 24 hours with or without acute ischemic lesion (s) on brain computed tomography or with acute ischemic diffusion-weighted imaging lesion (s) on magnetic resonance imaging that correspond (s) to the clinical presentations.
were defined as having AIS. Any hospitalization in the claims data that could be successfully linked to one of the hospitalizations in the registry data was regarded as a true episode of AIS. For hospitalizations in the claims data that could not be linked to any of the hospitalizations in the registry data, electronic medical records and brain images were reviewed by an experienced stroke neurologist (MTH) to determine whether they were true episodes of AIS. For hospitalizations determined not to be true episodes of AIS, the reasons were further categorized into the following groups: (1) subacute ischemic stroke (presenting between 11 and 30 days after symptom onset); (2) chronic ischemic stroke with known onset time (presenting more than 30 days after symptom onset, eg, hospitalization for inpatient rehabilitation); (3) chronic ischemic stroke with unknown onset time (chronic ischemic lesions identified on brain images); (4) brain edema due to subarachnoid hemorrhage, subdural hemorrhage, or intracerebral hemorrhage; (5) a tentative diagnosis of AIS, which was later excluded after clinical evaluation and imaging studies; and (6) other diagnoses (eg, encephalopathy, consciousness disturbance due to systemic diseases).

**Data Analyses**

We explored various case definition algorithms for identifying cases of AIS including (1) ICD-10-CM code of I63* as the primary diagnosis; (2) I63* as the primary diagnosis or first secondary diagnosis; (3) I63* as the primary, first secondary, or second secondary diagnosis; (4) I63* in any position of the diagnosis. PPV and sensitivity were calculated and the 95% confidence intervals (CIs) for binomial proportions were obtained using the Clopper-Pearson exact method. In a sensitivity analysis, we examined whether patient age would affect the validity of ICD-10-
CM coding in the claims data. We compared the PPVs between the elderly (aged 65 years or older) and young subgroups using the chi-square test. A two-tailed p value of <0.05 was considered statistically significant. Data analyses were performed using SPSS 25 (IBM SPSS Inc, Chicago, Illinois).

**Results**

From Jan 2018 to Dec 2019, a total of 1235 distinct hospitalizations due to AIS were identified from the stroke registry. During the same period, a total of 1382 distinct hospitalizations with an ICD-10-CM code of I63* in any position of the discharge diagnoses were identified from the inpatient claims data. After linking records from both data sources (Figure 1), 1227 hospitalizations were successfully linked and considered to be true positive episodes of AIS. Among the remaining 155 hospitalizations in the claims data that could not be linked to any record in the stroke registry, 54 were considered true positive episodes of AIS whereas 101 were determined to be false-positive episodes of AIS after manual review by the stroke neurologist. In addition, eight hospitalizations from the stroke registry were not identified in the claims data (Figure 1) and were thus determined to be false-negative episodes of AIS. As a result, the PPV and sensitivity of ICD-10-CM diagnosis of AIS were 92.7% (95% CI, 91.2% to 94.0%) and 99.4% (95% CI, 98.8% to 99.7%).

The reasons for the false-positive episodes of AIS are listed in Tables 1 and 2. The most common reason was chronic ischemic stroke with unknown onset time, followed by AIS as a tentative diagnosis but excluded by studies. Of the eight false-negative episodes of AIS, two were miscoded as I61 (nontraumatic intracerebral hemorrhage) and one as I62 (other and unspecified nontraumatic intracranial hemorrhage). The remaining five hospitalizations did not contain any diagnosis codes related to stroke (from I60 to I69).

Table 2 presents the final diagnoses of the 11 false-positive AIS cases. The table shows the number of cases for each diagnosis. The highest PPV was 99.8% for Subarachnoid hemorrhage, followed by Subdural hemorrhage and Intracranial hemorrhage or intraventricular hemorrhage with PPVs of 97.1% and 99.4%, respectively. The sensitivity of these diagnoses was also high, ranging from 94.0% to 99.8%.

**Table 2 Final Diagnoses of Miscoded Cases (N = 11)**

| Diagnosis                                         | n |
|---------------------------------------------------|---|
| Subarachnoid hemorrhage                           | 2 |
| Subdural hemorrhage                               | 1 |
| Intracranial hemorrhage or intraventricular hemorrhage | 2 |
| Brain metastasis                                  | 1 |
| Posterior reversible encephalopathy syndrome      | 1 |
| Hypoxic encephalopathy                            | 1 |
| Amyloid angiopathy                                | 1 |
| Central nervous system infection                  | 2 |

Table 3 gives the validity of various algorithms for identifying cases of AIS. When ICD-10-CM code of I63* as the primary diagnosis was used to identify AIS, the PPV was 99.8%. However, this algorithm would miss a substantial proportion of AIS hospitalizations with a sensitivity of 87.2% because conditions other than AIS were coded as the primary diagnosis. Table 4 summarizes the reasons for the 165 missed cases. Among them, the most common reason was infection or inflammation. In addition, 35 (21.2%) of these cases had prolonged hospitalization over 30 days. It is of note that when AIS was defined to be I63* in the first three positions of discharge diagnoses, the PPV of AIS increased to 97.1% with an acceptable loss of sensitivity from 99.4% to 97.4%.

In the sensitivity analysis that examined the effect of patient age on the validity of ICD-10-CM coding, the PPVs were 91.3% (813/890; 95% CI, 89.3% to 93.1%) in the elderly subgroup and 95.1% (468/492; 95% CI, 92.8% to 96.8%) in the young subgroup, respectively, with a statistically significant difference (p = 0.010).

**Discussion**

This study found that using ICD-10-CM code of I63* in any position of the discharge diagnoses to identify AIS in hospitalized patients yielded a PPV and sensitivity of 92.7% and 99.4%, respectively. The PPV of AIS increased to nearly 100% when AIS was restricted to those with I63* as the primary diagnosis, but at the cost of a decrease in the sensitivity of more than 12%. By contrast, when I63* was listed as the primary, first secondary, or second secondary diagnosis, both a PPV and sensitivity of greater than 97% could be achieved.

The study findings have some potential implications for researchers who desire to use the NHI claims database to conduct stroke studies in the future. First, the algorithm that uses I63* as the primary diagnosis to identify AIS will yield the highest PPV, thus ensuring that the
Table 3 The Validity of Various Algorithms for Identifying Cases of Acute Ischemic Stroke

| Algorithm | TP | FP | FN | PPV (95% CI), % | Sensitivity (95% CI), % |
|-----------|----|----|----|----------------|------------------------|
| I63\textsuperscript{a} as primary diagnosis | 1124 | 2 | 165 | 99.8 (99.4–100) | 87.2 (85.3–89.0) |
| I63\textsuperscript{a} as primary or first secondary diagnosis | 1201 | 20 | 88 | 98.4 (97.5–99.0) | 93.2 (91.7–94.5) |
| I63\textsuperscript{a} as primary, first secondary, or second secondary diagnosis | 1255 | 37 | 34 | 97.1 (96.1–98.0) | 97.4 (96.3–98.2) |
| I63\textsuperscript{a} in any position of diagnosis | 1281 | 101 | 8 | 92.7 (91.2–94.0) | 99.4 (98.8–99.7) |

Note: I63\textsuperscript{a} means all the diagnosis codes that begin with I63.

Abbreviations: CI, confidence interval; FN, false negative; FP, false positive; PPV, positive predictive value; TP, true positive.

Table 4 Reasons for False-Negative Episodes of AIS When I63\textsuperscript{a} as the Primary Diagnosis Was Used to Identify AIS (N = 165)

| Reason | n (%) |
|--------|-------|
| Infection or inflammation | 46 (27.9) |
| Neurological symptoms\textsuperscript{a} | 22 (13.3) |
| Malignancy | 13 (7.9) |
| Heart diseases | 13 (7.9) |
| Ventilator dependency | 11 (6.7) |
| Traumatic brain injury | 10 (6.1) |
| Concomitant cerebrovascular disease other than AIS | 9 (5.5) |
| Renal diseases | 6 (3.6) |
| Diabetes mellitus | 5 (3.0) |
| Aortic aneurysm or dissection | 4 (2.4) |
| Gastrointestinal bleeding | 4 (2.4) |
| Miscellaneous\textsuperscript{b} | 22 (13.3) |

Notes: \textsuperscript{a}Neurological symptoms coded as the primary diagnosis, including consciousness change, numbness, and visual loss; \textsuperscript{b}Miscellaneous diagnoses including hypertensive disease (n=2), hernia (n=2), ischemic bowel (n=2), arteriovenous fistula (n=1), bile duct stone (n=1), burn injury (n=1), depression (n=1), dermatitis (n=1), bone fracture (n=1), herniated intervertebral disc (n=1), hypercalcemia (n=1), hyperlipidemia (n=1), hyponatremia (n=1), joint dislocation (n=1), liver disease (n=1), neonatal hypoxia (n=1), pancytopenia (n=1), radiculopathy (n=1) retinal artery occlusion (n=1).

Abbreviation: AIS, acute ischemic stroke.

Compared with a previous study that validated AIS diagnosis using ICD-9-CM codes,\textsuperscript{10} the PPVs of ICD-10-CM codes (I63\textsuperscript{a}) for AIS were higher than those of ICD-9-CM codes (433\textsuperscript{a} and 434\textsuperscript{a}). This finding is not unusual given that the use of brain imaging studies, such as computed tomography and magnetic resonance imaging, has increased over time. As observed in a systematic review, the PPVs of diagnostic codes for stroke has improved over the past decades.\textsuperscript{14} A more recent study also found that in a cohort of US patients, using I63\textsuperscript{a} as the primary discharge diagnosis to identify patients admitted with AIS yielded a high PPV of 97.6%, which is equal to or even higher than the PPV values when ICD-9-CM codes were used for AIS.\textsuperscript{15} However, the US cohort was limited to patients coded for AIS in the primary discharge diagnosis, making it impossible to assess other performance metrics such as sensitivity. Other possible explanations could include the influence of the nationwide quality improvement initiative on the quality of stroke care\textsuperscript{16} and a newly implemented pay-for-performance program for AIS in Taiwan. Similar to the pay-for-performance program for diabetes care,\textsuperscript{17,18} this program uses ICD codes in the primary diagnosis field to define AIS. These quality improvement activities may provide an incentive for hospitals to improve the accuracy of diagnostic coding for AIS because a high false-positive rate of diagnosis of AIS may create a bias for the measurement of hospital performance.

Regarding the false-positive cases incorrectly coded as AIS, nearly half of them were determined to have a chronic ischemic stroke with unknown onset time and one fourth were tentatively diagnosed with AIS but received an alternative diagnosis after studies. The false-positive rate was even higher in patients aged 65 years or older. It may reflect the difficulty of diagnosing AIS in elderly patients with multiple comorbidities and pre-existing disability. The liberal use of imaging studies may be helpful in this situation. Nevertheless, sensitive imaging tools may even unmask chronic and silent.
ischemic lesions, thus increasing the difficulty in diagnosis and decision-making for further treatment. The difference in PPVs of ICD coding for AIS between elderly and young subjects should be kept in mind when conducting claims data research.

This study has several limitations worth mentioning. First, this is a single-center study. It is unclear whether the study results can be generalized to the population covered by Taiwan’s NHI as well as stroke populations in other countries. Further multicenter validation studies are warranted to confirm or compare these findings. Second, the ascertainment of AIS was based on the data from the stroke registry. Even though registry databases were generally used as the reference sources in similar studies and the registry data were collected by trained study nurses following a standard protocol, whether the registry data can be considered the gold standard is open to questions. Ideally, a cross-sectional study is required to corroborate the accuracy of the stroke registry before it can be used as a reference standard. Third, using the current study design, we were unable to discover patients with acute ischemic stroke who had been missed by both the stroke registry and ICD codes (false negatives). Furthermore, we did not have any data regarding the number of patients who did not have acute ischemic stroke and were not coded as such (true negatives). Therefore, we were unable to calculate specificity and the negative predictive value. Fourth, hospitalizations that could not be linked to those in the registry data were manually reviewed by only one neurologist. Nevertheless, these hospitalizations accounted for only 11% (155/1382) of the hospitalizations from the claims data. Fifth, although including ICD-10 code I64 (acute but ill-defined stroke) to identify AIS might help capture more cases with AIS with little decrease in the PPV, the validity of I64 could not be examined because I64 does not exist in ICD-10-CM.

Conclusion
This validation study demonstrated the validity of various case definition algorithms for AIS in Taiwan’s NHI claims database based on ICD-10-CM coding. The study results can provide a reference for future researchers aiming to conduct stroke research using this claims database.

Acknowledgments
This research was supported by the E-Da Hospital (grant number EDAHP106047). 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. The authors would like to thank Mr Hsing-Fen Lin for the assistance of statistical analysis during the completion of the manuscript and Ms Li-Ying Sung for English language editing.

Disclosure
The authors report no conflicts of interest in this work.

References
1. Johnson W, Onuma O, Owolabi M, Sachdev S. Stroke: a global response is needed. Bull World Health Organ. 2016;94(9):634–634. doi:10.2471/BLT.16.181636
2. Feigin VL, Forouzanfar MH, Krishnamurthi R, et al. Global and regional burden of stroke during 1990–2010: findings from the global burden of disease study 2010. Lancet. 2014;383(9913):245–254. doi:10.1016/s0140-6736(13)61953-4
3. GBD 2016 Lifetime Risk of Stroke Collaborators. Global, regional, and country-specific lifetime risks of stroke, 1990 and 2016. N Engl J Med. 2018;379(25):2429–2437. doi:10.1056/NEJMoal804492.
4. Yu AYX, Holodinsky JK, Zerna C, et al. Use and utility of administrative health data for stroke research and surveillance. Stroke. 2016;47(7):1946–1952. doi:10.1161/STROKEAHA.116.012390
5. Ung D, Kim J, Thrift AG, et al. Promising use of big data to increase the efficiency and comprehensiveness of stroke outcomes research. Stroke. 2019;50(5):1302–1309. doi:10.1161/STROKEAHA.118.020372
6. Sung S-F, Hsieh C-Y, Hu Y-H. 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
7. Jollis JG. Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. Ann Intern Med. 1993;119(8):844–850. doi:10.7326/0003-4819-119-8-199310150-00011
8. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. Int J Epidemiol. 1996;25(2):435–442. doi:10.1093/ije/25.2.435
9. Cheng C-L, Kao Y-HY, Lin S-J, Lee C-H, Lai M-L. Validation of the national health insurance research database with ischemic stroke cases in Taiwan. Pharmacoeconom Drug Saf. 2011;20(3):236–242. doi:10.1002/pds.2087
10. Hsieh C-Y, Chen C-H, Li C-Y, Lai M-L. Validating the diagnosis of acute ischemic stroke in a national health insurance claims database. J Formos Med Assoc. 2015;114(3):254–259. doi:10.1016/j.jfma.2013.09.009
11. Sung S-F, Hsieh C-Y, Lin H-J, Chen Y-W, Yang Y-HK, Li C-Y. Validation of algorithms to identify stroke risk factors in patients with acute ischemic stroke, transient ischemic attack, or intracerebral hemorrhage in an administrative claims database. Int J Cardiol. 2016;215:277–282. doi:10.1016/j.ijcard.2016.04.069
12. Hsieh C-Y, Su -C-C, Shao S-C, et al. Taiwan’s national health insurance research database: past and future. Clin Epidemiol. 2019;11:349–358. doi:10.2147/CEP.S196293
13. Hsieh F-I, Lien L-M, Chen S-T, et al. Get with the guidelines-stroke performance indicators: surveillance of stroke care in the Taiwan stroke registry: get with the guidelines-stroke in Taiwan. Circulation. 2010;122(11):1116–1123. doi:10.1161/CIRCULATIONAHA.110.936526
14. McCormick N, Bhole V, Lacaille D, Avina-Zubieta JA. Validity of diagnostic codes for acute stroke in administrative databases: a systematic review. PLoS One. 2015;10(8):e0135834. doi:10.1371/journal.pone.0135834
15. Alhajji M, Kawsara A, Alkhouli M. Validation of acute ischemic stroke codes using the international classification of diseases tenth revision. Am J Cardiol. 2020;125(7):1135. doi:10.1016/j.amjcard.2020.01.004

16. Hsieh F-I, Jeng J-S, Chern C-M, et al. Quality improvement in acute ischemic stroke care in Taiwan: the breakthrough collaborative in stroke. Wang X, ed. PLoS One. 2016;11(8):e0160426. doi:10.1371/journal.pone.0160426

17. Chou C-W, Kung P-T, Chou W-Y, Tsai W-C. Pay-for-performance programmes reduce stroke risks in patients with type 2 diabetes: a national cohort study. BMJ Open. 2019;9(10):e026626. doi:10.1136/bmjopen-2018-026626

18. Lee I-T, Hsu -C-C, Sheu W-H, Su S-L, Wu Y-L, Lin S-Y. Pay-for-performance for shared care of diabetes in Taiwan. J Formos Med Assoc. 2019;118(Suppl 2):S122–S129. doi:10.1016/j.jfma.2019.08.011

19. Meinel TR, Kaesmacher J, Roten L, Fischer U. Covert brain infarction: towards precision medicine in research, diagnosis, and therapy for a silent pandemic. Stroke. 2020;51(8):2597–2606. doi:10.1161/STROKEAHA.120.030686

20. Kim JY, Lee K-J, Kang J, et al. Development of stroke identification algorithm for claims data using the multicenter stroke registry database. PLoS One. 2020;15(2):e0228997. doi:10.1371/journal.pone.0228997