








Identifying long-term conditions in New Zealand general practice using structured and unstructured data: a cross-sectional study

Younhyang Catherine Choi ¹, Katrina Poppe ², Vanessa Selak ¹, Allan Ronald Moffitt ³, Claris Yee Seung Chung ⁴, Jane Ullmer ⁵, Sue Wells ⁶

To cite: Choi YC, Poppe K, Selak V, *et al.* Identifying long-term conditions in New Zealand general practice using structured and unstructured data: a cross-sectional study. *BMJ Health Care Inform* 2025;**32**:e101393. doi:10.1136/bmjhci-2024-101393

Received 05 December 2024
Accepted 27 April 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

¹Department of Epidemiology and Biostatistics, University of Auckland, Auckland, New Zealand

²Department of Medicine, University of Auckland, Auckland, New Zealand

³Procare Health Limited, Auckland, New Zealand

⁴Department of Accounting and Information Systems, University of Canterbury, Christchurch, New Zealand

⁵School of Population Health, The University of Auckland Faculty of Medical and Health Sciences, Auckland, New Zealand

⁶Department General Practice and Primary Healthcare, University of Auckland, Auckland, New Zealand

Correspondence to

Younhyang Catherine Choi;
younhyang.choi@auckland.ac.nz

ABSTRACT

Objectives This study examined whether incorporating free-text entries into structured general practice records improves the detection of long-term conditions (LTCs) and multimorbidity (MM) in New Zealand (NZ) general practices.

Methods Data from 374 071 deidentified individuals in general practices were analysed to identify 61 LTCs. Structured data were extracted using Read codes from a national master list, and clinical raters independently identified condition-related free-text, including synonyms, negation terms and common misspellings in randomised samples. Keywords were categorised and refined through ten iterative tests. Programmatic text classification was developed and assessed against gold-standard clinician ratings, using sensitivity, specificity, positive predictive value (PPV) and F_1 -score.

Results A quarter of general practitioner classifications contained either unrecognised Read codes or consisted of free-text only. Clinician inter-rater reliability was high ($\kappa \geq 0.9$). Compared with clinical gold standard, text classification yielded an average sensitivity of 88%, specificity of 99% and PPV of 95%, with an F_1 -score range of 82%–95%. Incorporating free text increased LTC prevalence from 42.1% to 46.3%, reducing misclassification of MM diagnoses by identifying 12 626 additional patients with MM and 15 972 additional patients with at least one LTC.

Discussion In the course of workflow, general practitioners face barriers to accurate LTC coding or may simply annotate with text-based descriptions. Programmatic text classification has demonstrated high performance and identified many more patients receiving LTC care.

Conclusions Combining structured and unstructured data optimises MM detection in NZ general practices and has the potential to improve case management, follow-up care and allocation of healthcare resources.

INTRODUCTION

Multimorbidity (MM), the presence of two or more long-term conditions in an individual, is associated with reduced quality of life,¹ increased healthcare needs, hospitalisations

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Accurate identification of multimorbidity (MM), defined as the presence of two or more long-term conditions, is essential for effective patient management. In New Zealand (NZ), approximately 25% of the population is estimated to have MM based on the M3 MM index derived from routinely collected national hospitalisation data; however, this approach likely underestimates the true community burden.
- ⇒ Augmenting hospitalisation records with general practice data has been proposed to provide a more accurate picture of MM. In NZ general practices, long-term condition data are recorded in both structured (eg, Read codes) and unstructured free-text entries, the latter presenting significant classification challenges.

WHAT THIS STUDY ADDS

- ⇒ This study presents a robust methodology that combines the extraction of structured Read codes with the analysis of unstructured free-text data from general practice records, thereby enhancing the identification of patients with multiple long-term conditions.
- ⇒ Around a quarter of general practitioner classifications have codes that are not recognised in a national master Read code list or are text-only entries. However, the procedure effectively converts their textual data into analysable data elements to identify MM conditions.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ Using Read coded classifications alone underestimates long-term condition prevalence and complexity of care. The ability to leverage unstructured text-based records provides a more accurate and comprehensive understanding of morbidity in general practice. It will enable targeted care for those not otherwise identified as having MM, support optimal health service planning and funding, and better inform workforce development needs.

and premature mortality.^{2–4} In 2017, an MM index, M3, was developed and validated for 1-year mortality risk from all adult New Zealand (NZ) residents using 61 long-term condition categories from routine hospital admission data.⁵

This index estimated that MM affected approximately 8% of the NZ general population, with around 91% of affected individuals aged 65 years or older.² International evidence indicates a strong link between socioeconomic factors and MM. For instance, systematic reviews indicate that lower education is associated with a roughly 64% increased risk of MM, while individuals in the lowest income bracket are over four times more likely to develop MM compared with those in the highest income bracket.⁶ Moreover, findings from Scotland, based on primary care records, demonstrate that the onset of MM occurs 10–15 years earlier in the most deprived populations, with deprivation particularly linked to MM involving mental health disorders.⁷ However, because the M3 index relies on International Classification of Disease 10th edition Australian Modification (ICD-10-AM) diagnostic coding following hospital admissions, it under-represents long-term conditions that are primarily managed in general practice settings such as dietary controlled diabetes mellitus, gout, dementia or angina.⁸

For this study, ‘general practice data’ refer to the electronic health records (EHRs) maintained by NZ general practices. These records capture comprehensive patient information through both structured coding systems (mainly Read codes⁹) and unstructured free-text entries. Recent analyses using general practice EHRs have highlighted these limitations. For example, a study of 454 367 general practice patients in NZ¹⁰ found that 24% of individuals diagnosed with ischaemic cardiovascular disease by their general practitioners (GPs) had had no corresponding hospital admission. Similarly, data from the QRISK3 GP database in the UK revealed that 23% of angina and 55% of transient ischaemic attack outcomes were solely documented in GP records.¹¹ While extraction of structured fields can be performed accurately and consistently, manual extraction from unstructured free-text is labour-intensive, prone to human error and presents significant challenges, especially in large-scale studies of diverse chronic diseases.^{12–14}

We aimed to develop a robust procedure for extracting both structured and unstructured data from general practice EHRs in NZ. This approach will facilitate research comparing MM representation between secondary and general practice settings, quantifying MM in the community and investigating the association of MM with cardiovascular disease outcomes.

METHODS

We investigated the classification of long-term conditions in a large cohort of general practice patients. This involved identifying relevant Read codes (the clinical coding system used in NZ general practice to document

patient diagnoses and procedures) for each ICD-10-AM diagnostic code within the 61 long-term condition categories of the M3 index.⁵ Additionally, we developed a programmable text-based classification procedure to extract conditions associated with non-standard Read codes and free-text documented in general practice EHRs.

Data

General practice data were obtained from ProCare Health, a primary health organisation (PHO) based in Auckland, NZ, which has around 170 member practices and represents 51% of Auckland’s population. The data were extracted from multiple general practices under the ProCare PHO, with all practices contributing to the dataset. As of 1 January 2014, 623 475 patients aged 18 years and over were enrolled. Depending on the practice electronic management system and with practice approval, ProCare currently extracts the EHRs of enrolled patients to a secure central information repository, to support population health gain and quality improvement across the network. Following national ethics and ProCare clinical governance approval, deidentified individual patient data from people aged 18 years and over were used for this study. The dataset included both structured (Read coded) and unstructured (free-text) documentation of long-term conditions and date of entry, directly extracted from the EHRs entered by GPs.

To address data completeness, this study considered only those patients whose data were eligible for extraction from the patient management systems (PMS). Patients who were not eligible due to technical challenges associated with varying PMS and differing coding systems (SNOMED vs Read codes) were excluded from the analysis, resulting in a reduction from 623 475 to 374 071 patients. Since this exclusion was based on eligibility criteria rather than missing data, no imputation methods were applied.

The Read code reference list, available from ProCare PHO and accessible on the Accident Compensation Corporation website,⁹ was reviewed, and the conditions were matched with each of the 61 categories in the M3 index. Clinical data on long-term conditions relevant to the M3 index were categorised into four groups:

- ▶ Group 1: Structured data—valid Read codes plus Read code text definitions.
- ▶ Group 2: Unstructured data—partially entered or invalid Read codes plus descriptions.
- ▶ Group 3: Unstructured data—free-text notes only.
- ▶ Group 4: No data entered.

Records in group 4, which accounted for 0.4% of the total records, contained no clinical recording of any medical conditions or free-text notes and were thus excluded from the rest of the analysis.

Establishing keywords

500 records were randomly selected from each of groups 1, 2 and 3, creating a sample of 1500 entries recorded

in patient medical history classification lists. The records were independently reviewed by two clinical assessors (SW and KP), who identified keywords relating to M3 condition categories in each entry, their potential synonyms, spelling errors, abbreviations and negation words (such as 'no evidence', 'resolved' or 'family history'). The assessors compared their selections, and any disagreement was resolved through consensus. Another 1500 entries were then extracted, and the process was repeated with 4500 entries assessed in this way. Inter-rater agreement was evaluated using Cohen's kappa after each iteration of 1500 entries.

Implementing the categorisation strategy

To preprocess data, (a) all text entries were converted to lowercase to enable case-insensitive searches; (b) uninformative entries such as 'not applicable' or 'unknown' were removed as unnecessary whitespace; (c) special characters were removed, except for question marks, a minus sign (ie, -ve representing 'negative'), and the pound sign (#, representing a fracture). Numbers within the text were retained.

Excluding the 4500 records that had already undergone keyword detection, a random sample of 1000 preprocessed text records from groups 1, 2 and 3 was taken (figure 1). Code was developed by YCC in R statistical software to perform computer-generated classification of M3 condition categories based on entry terms and negation rules. This programmatic classification procedure will be referred to as the 'R_M3Text_Classification' in this and future publications.

The accuracy of the classification of free-text entries into M3 categories was assessed by comparing its output with manual classification performed by a clinical expert (SW). In each iteration, the classification process involved evaluating how well the identified keywords and negation phrases corresponded to the M3 categories, with attention to cases where anatomical context was relevant. For example, entries with 'aneurysm' were initially categorised under 'aortic and other aneurysms' (eg, dissection of aortic aneurysm), but if they were combined with 'cerebral', they were reclassified under 'cerebrovascular disease' (eg, cerebral artery aneurysm). This

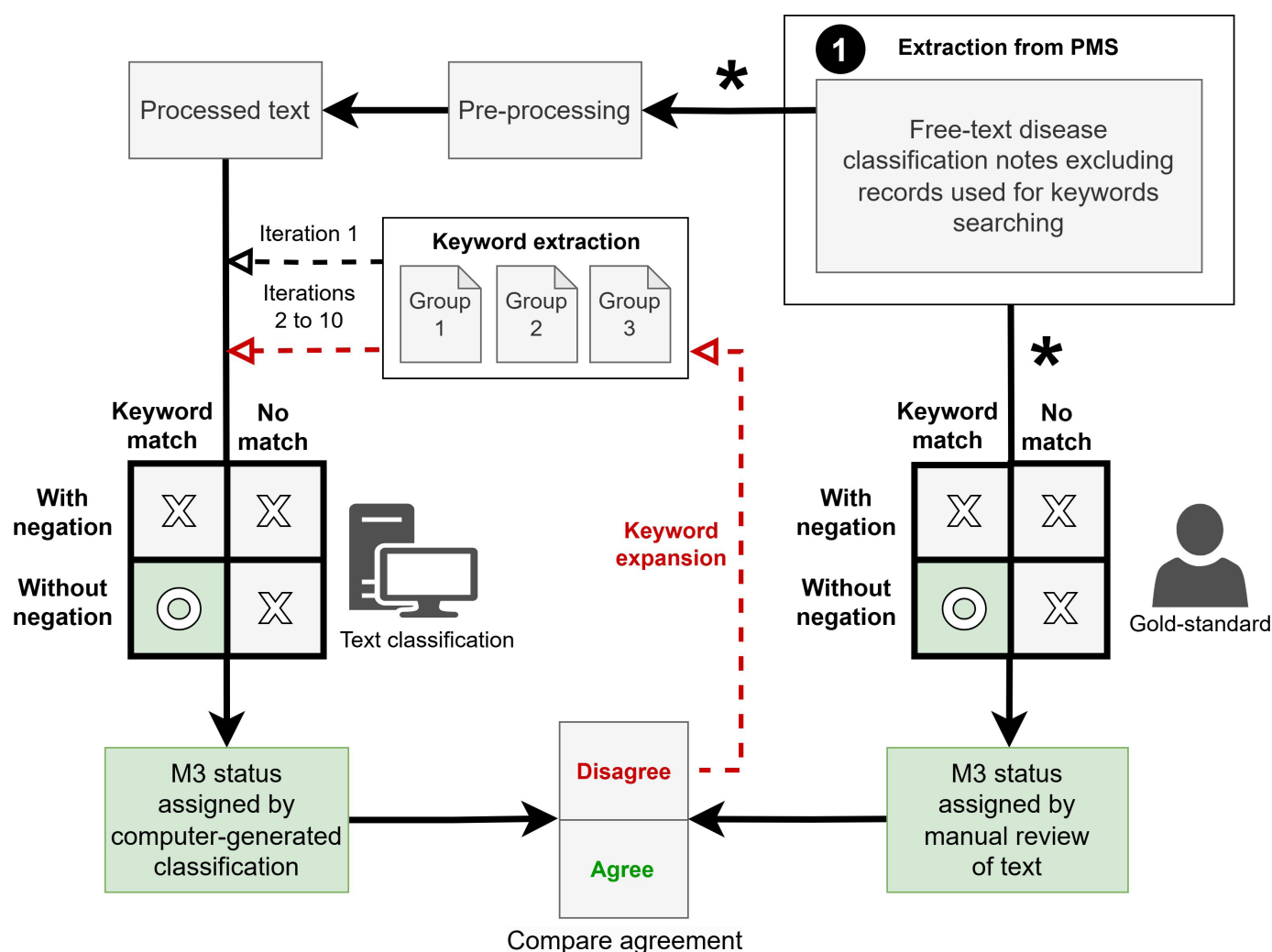


Figure 1 Text-data mining and validation processes (*Information extracted from the patient management systems (PMS) underwent simultaneous review and classification by both software and a clinical expert).

review process was iterated 10 times, with a new random sample of 1000 entries reviewed in each iteration. As new keywords, synonyms, abbreviations, spelling mistakes and negation phrases were identified, they were incorporated into the classification model. Any questions or uncertainties about the meaning of a text entry or corresponding M3 categorisation were resolved by clinical consensus (SW and KP).

To evaluate the performance of the 'R_M3Text_Classification' against manual classifications, we used standard metrics after each iteration: sensitivity (true positive rate), specificity (true negative rate), positive predictive value (percentage of positive results that are true positive) and F_1 -score. The F_1 -score represents the harmonic mean of positive predictive value and sensitivity and serves as an indicator of result quality, with 1 indicating optimal performance and 0 indicating poor performance. We aimed to optimise sensitivity without compromising specificity. All methodologies were performed using R Statistical Software V.4.3.1 (the R codes are available on request).

- Sensitivity: $TP/(TP+FN)$, where TP is true positive, FN is false negative.
- Specificity: $TN/(TN+FP)$, where TN is true negative, FP is false positive.
- Precision (or positive predictive value): $TP/(TP+FP)$.
- F_1 -score: $(2 \times \text{precision} \times \text{sensitivity}) / (\text{precision} + \text{sensitivity})$.

Detecting negation in clinical notes

Clinical notes often include symptoms that are absent or conditions that can be ruled out. Terms such as 'no', 'without', 'never had' and 'no signs or symptoms of' were used to categorise entries as negated and not to be classified under certain M3 conditions (eg, 'never had asthma', 'mammogram no evidence of breast cancer'). To improve classification coding, phrases containing these negation terms were reviewed to identify common patterns.

RESULTS

A total of 623 475 patients, aged ≥ 18 years and enrolled in ProCare PHO as of 1 January 2014, were included in the study (median age 45 years (IQI 32–58); 53% female). Following the patient management extraction process, this number was reduced to 374 071 patients (60% of the original cohort; median age 46 years (IQI 33–58); 55% female), due to technical challenges related to the variability of PMS and differing coding systems. Deidentified individual patient data from the disease classification portion of the PMS were investigated, yielding a total of 7 154 762 PMS entries (an average of 19 entries per person). These records accounted for 99.6% of the total records and were characterised by the presence of Read codes, non-standard Read coding and/or free-text notes in the disease classification fields (figure 2).

Of the 7 154 762 entries with information in the disease classification portion of the clinical record, 24.1% ($n=1\,726\,168$) were identified as either non-standard or

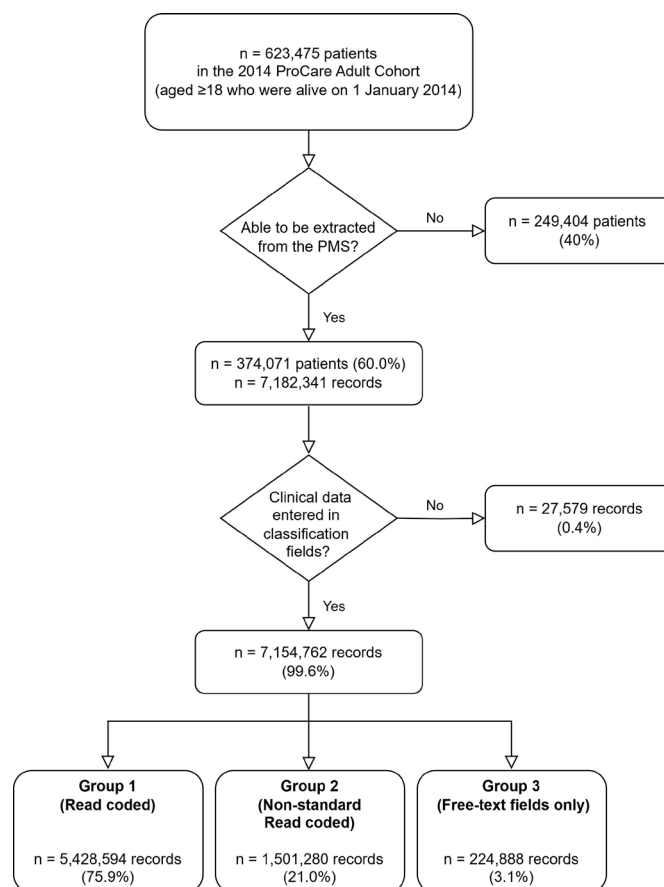


Figure 2 Dataset preparation. PMS, patient management system.

incompletely coded or stored as unstructured free-text notes (groups 2 and 3).

There were 224 888 records that were only unstructured free-text, with each sentence containing a maximum of 25 words and a median of 3 words (IQI 1–4). Cohen's kappa for the inter-rater reliability of assessing keywords was 0.95 (95% CI 0.92 to 0.97), 0.90 (95% CI 0.87 to 0.93) and 0.99 (95% CI 0.98 to 1.00) for groups 1, 2 and 3, respectively, indicating a high level of agreement among the assessors' judgements.

Assessment of negation terms revealed five dominant patterns: (1) preceding terms that appear before disease findings, (2) following terms that appear after disease findings, (3) pseudo-negation terms that indicate negation but actually represent double negatives (eg, not ruled out), (4) ambiguous phrasing (eg, possibly, unlikely) and (5) specific terms that only act as negation indicators for particular disease conditions. An example in figure 3 shows sentences that may be used to describe colitis as a bowel disease, including typical negation terms.

The 'R_M3Text_Classification' procedure provided the sensitivity, specificity, positive predictive value and F_1 -score for each testing iteration (table 1). The sensitivity of identifying patients with one or more long-term conditions was lowest in the first iteration at 72.6% (95% CI 64.6% to 79.7%) and improved with all subsequent iterations (mean 90.2% (95%

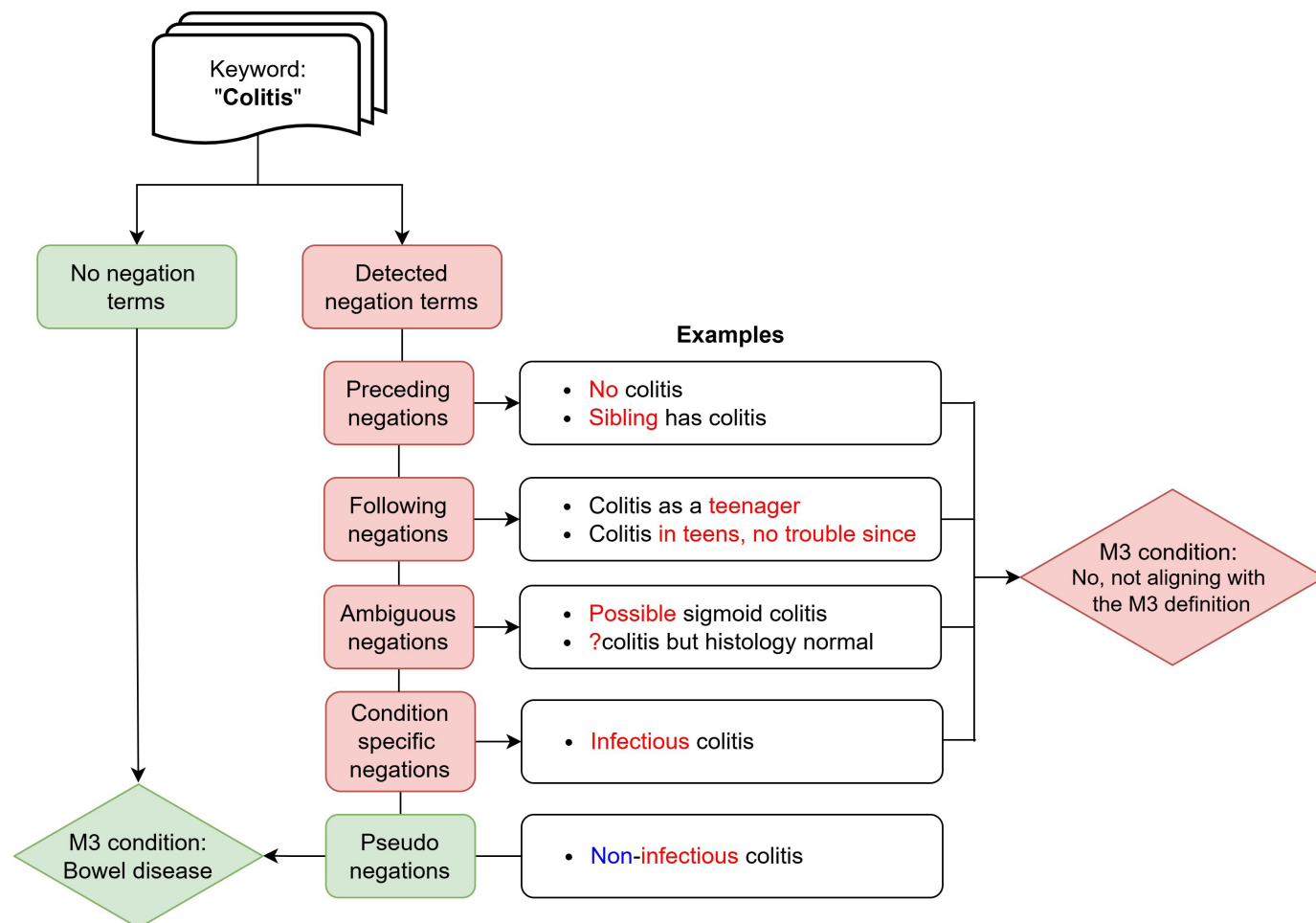


Figure 3 Negation detection in disease classification.

CI 88.6% to 91.9%)). Specificity ranged from 98.6% (95% CI 97.5% to 99.3%) to 100% (95% CI 99.5% to 100%), consistently surpassing the sensitivity in all iterations. The positive predictive value ranged from 92.3% (95% CI 86.6% to 96.1%) to 100% (95% CI 98.0% to 100%) and F₁-score from 81.5% to 94.7%.

The primary reason for incorrect classifications was simple spelling errors in free-text notes.

The prevalence of 61 M3 conditions in the cohort of 374 071 patients was 42.1% based solely on Read coded data. After incorporating the text-based information using the procedure developed in this study,

Table 1 Performance measurement results

	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive Predictive value, % (95% CI)	F ₁ -score, %
Iteration 1	72.6 (64.6 to 79.7)	99.1 (98.2 to 99.6)	93.0 (86.6 to 96.9)	81.5
Iteration 2	90.3 (84.3 to 94.6)	98.7 (97.7 to 99.3)	92.3 (86.6 to 96.1)	91.3
Iteration 3	86.7 (80.5 to 91.5)	98.9 (97.9 to 99.5)	94.1 (89.1 to 97.3)	90.2
Iteration 4	93.5 (88.4 to 96.8)	98.8 (97.8 to 99.4)	93.5 (88.4 to 96.8)	93.5
Iteration 5	92.8 (88.3 to 96.0)	99.0 (98.0 to 99.6)	95.8 (91.8 to 98.2)	94.3
Iteration 6	91.6 (85.8 to 95.6)	99.4 (98.6 to 99.8)	96.3 (91.6 to 98.8)	93.9
Iteration 7	89.9 (84.3 to 94.0)	98.6 (97.5 to 99.3)	92.6 (87.5 to 96.1)	91.2
Iteration 8	87.7 (82.0 to 92.1)	99.6 (98.9 to 99.9)	98.1 (94.6 to 99.6)	92.6
Iteration 9	89.8 (84.6 to 93.8)	98.6 (97.6 to 99.3)	93.9 (89.3 to 96.9)	91.8
Iteration 10	89.9 (84.9 to 93.6)	100 (99.5 to 100)	100 (98.0 to 100)	94.7

the prevalence of M3 conditions increased to 46.3%. Not considering text-based classifications would misclassify 15 972 patients as not having an M3 long-term condition category and 12 626 patients as having no MM.

DISCUSSION

This study highlights the importance of incorporating unstructured free-text data from general practice EHRs to provide a more accurate representation of the burden of long-term conditions. Our analysis, which represents the first large multipractice NZ study of MM in general practice, revealed a substantial number of multimorbid patients whose conditions were not captured in Read coded data. By augmenting structured data with free-text analysis, our 'R_M3Text_Classification' procedure improved the identification of M3 index conditions, achieving very high specificity, sensitivity, positive predictive value and F₁-score.

The sensitivity of the first iteration was 72.6%. This initial value was primarily due to over detection of conditions that were acute or short term, related to family history, or resulted from routine screening. For example, people with gestational diabetes, a family history of gynaecological cancer, negative HIV or breast cancer screening tests and previous benign breast biopsy results were often misclassified as having a long-term condition. To improve sensitivity while minimising false positives, the programme was refined to accurately identify when keywords indicated conditions were not long term. The highly variable nature of free-text, particularly in a database sourced from multiple clinicians, meant that it was not possible to identify all permutations of phrasing. Despite these challenges, our methodology achieved consistently high specificity—minimising false positives—which is crucial for accurate condition identification in general practice.

Two clinical experts collaboratively developed key terms, resolved ambiguities and refined the classification rules, while our statistical expert translated those decisions into logical programmable steps. This iterative process has demonstrated that even a relatively straightforward text classification approach can yield robust performance without the resource-intensive requirements of advanced natural language processing (NLP) techniques.

Our findings align with similar studies that aimed to harness unstructured free-text data from EHRs to enhance disease classification and management. Recent research^{15–18} employing advanced NLP and machine learning techniques has successfully analysed unstructured patient medical records, enabling effective identification of patients with specific diseases. However, these methods often necessitate

extensive computational resources and large datasets for training, which can limit their applicability in resource-constrained settings. In contrast, our study demonstrated that a more straightforward text classification approach, such as the 'R_M3Text_Classification' procedure, can yield high performance metrics without the complexities of NLP techniques, particularly in general practice free-text data. In line with recent work by Hossain *et al.*,¹⁹ which highlighted challenges with misclassification due to ambiguous phrasing and negation in free-text entries, our study reinforces the notion that while sophisticated algorithms can enhance accuracy, simpler methods can also be effective when tailored to the specific characteristics of the data. These findings suggest that our approach may serve as a practical alternative for general practice settings, where clinician time and resources are often limited.

While our study yielded promising results, it is important to acknowledge limitations such as dependence on the quality of free-text records and lack of specific keywords for some conditions, particularly when the condition descriptions are broad. For example, it was difficult to identify keywords for venous insufficiency and uncomplicated hypertension that were specific enough to avoid false positives. Another limitation arises from the presence of pseudo-negations, such as double negatives (eg, not ruled out) or ambiguous negations, which can pose challenges for the procedure and potentially lead to misclassification.

While challenges remain, including dependency on the quality of free-text records and the difficulty of capturing all phrasing variations, our study provides a practical alternative for general practice settings. For example, consider a patient whose coded records indicate atrial fibrillation and coronary disease. However, on incorporating free-text entries from the same records, we identified that the patient had also developed heart failure—a condition not captured by structured data alone. This case exemplifies the broader implications of our approach, demonstrating that augmenting structured data with free-text analysis enhances MM detection and provides a more accurate representation of patients' health status. Studies such as Owen *et al.*²⁰ have shown that the temporal sequence of MM diagnoses can significantly affect life expectancy, underscoring the clinical importance of capturing the full disease burden for optimal decision-making. Integration of our procedure into the PMS could improve medical documentation precision, enhance clinical decision-making and ultimately support better patient outcomes and health policy decisions. External validation across diverse patient populations is needed to further assess the applicability and effectiveness of this approach. This would provide valuable insights into its functionality in different general practice settings.

CONCLUSIONS

Unstructured clinical data from NZ general practice records is a valuable addition to the identification of patients with multiple conditions that may not be readily apparent in structured data. Manually extracting information from these records is time-consuming and prone to human error. We have developed a programme classification procedure that efficiently and consistently extracts information from extensive clinical notes for 61 long-term conditions. The approach not only provides a more accurate identification and count of adults with MM but also facilitates future research on MM and outcomes by eliminating the need for manual review of free-text data. Future research can build on this methodology by exploring how the improved identification of MM affects treatment outcomes and long-term health trajectories.

X Allan Ronald Moffitt @Allan_Moffitt

Acknowledgements The authors express their gratitude to Dr Raina Elley for her valuable advice and insightful discussions on the study design.

Contributors SW, YCC and KP planned and designed the study. SW, KP and ARM were involved in the data collection process. YCC analysed the data with input from SW and KP. YCC, KP, VS, ARM, CYSC, JU and SW were involved in data interpretation. YCC, SW and KP drafted the manuscript and all authors revised the manuscript. All authors approved the final submitted version and agreed to be accountable for the manuscript. YCC is the guarantor. The corresponding author, YCC, confirms that she had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Funding This work was supported by the New Zealand Health Research Council (HRC), programme 20/304. KP is funded in part through a Heart Health Research Trust Senior Fellowship, which is supported by the National Heart Foundation of New Zealand under grant 1866. JU is funded by the NZ College of Public Health Medicine and the HRC Vascular Risk Equity for All New Zealanders programme under grant 21/712.

Disclaimer The funders of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit the report for publication.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study is part of a programme of research originally approved by the Northern Region Ethics Committee in 2003 (AKY/03/12/314), with subsequent approval by the National Multi Region Ethics Committee in 2007 (MEC07/19/EXP) and with annual reapproval since 2007 by the National Multi Region Ethics Committee as part of a vascular research programme (2023 EXP 18564). Individual patient consent was not required as all data are deidentified.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data from this manuscript can be requested with data access proposals. For the ProCare data, applications will only be granted and data access provided after agreement from the contributing provider and after ethical approval by the New Zealand Multi-Region Ethics Committee.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Yeunhyang Catherine Choi <http://orcid.org/0000-0002-2085-0673>

Katrina Poppe <http://orcid.org/0000-0002-4418-4476>
 Vanessa Selak <http://orcid.org/0000-0002-9824-8674>
 Allan Ronald Moffitt <http://orcid.org/0009-0008-6944-4036>
 Claris Yee Seung Chung <http://orcid.org/0000-0002-3931-466X>
 Jane Ullmer <http://orcid.org/0009-0003-4687-2699>
 Sue Wells <http://orcid.org/0000-0003-2942-9524>

REFERENCES

- 1 Stairmand J, Gurney J, Stanley J, *et al*. The impact of multimorbidity on people's lives: a cross-sectional survey. *The New Zealand Medical Journal (Online)* 2018;131:78–90.
- 2 Stanley J, Semper K, Millar E, *et al*. Epidemiology of multimorbidity in New Zealand: a cross-sectional study using national-level hospital and pharmaceutical data. *BMJ Open* 2018;8:e021689.
- 3 Kabir A, Tran A, Ansari S, *et al*. Impact of multimorbidity and complex multimorbidity on mortality among older Australians aged 45 years and over: a large population-based record linkage study. *BMJ Open* 2022;12:e060001.
- 4 Glynn LG, Valderas JM, Healy P, *et al*. The prevalence of multimorbidity in primary care and its effect on health care utilization and cost. *Fam Pract* 2011;28:516–23.
- 5 Stanley J, Sarfati D. The new measuring multimorbidity index predicted mortality better than Charlson and Elixhauser indices among the general population. *J Clin Epidemiol* 2017;92:99–110.
- 6 Skou ST, Mair FS, Fortin M, *et al*. Multimorbidity. *Nat Rev Dis Primers* 2022;8:48.
- 7 Barnett K, Mercer SW, Norbury M, *et al*. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet* 2012;380:37–43.
- 8 Gini R, Francesconi P, Mazzaglia G, *et al*. Chronic disease prevalence from Italian administrative databases in the VALORE project: a validation through comparison of population estimates with general practice databases and national survey. *BMC Public Health* 2013;13:1–11.
- 9 Accident Compensation Corporation. Read codes. Available: <https://www.acc.co.nz/for-providers/lodging-claims/read-codes> [Accessed 10 Aug 2022].
- 10 Wells S, Poppe KK, Selak V, *et al*. Is general practice identification of prior cardiovascular disease at the time of CVD risk assessment accurate and does it matter? How simple mistakes and short-term bias elevate cardiovascular risk. 2018;131.
- 11 Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;357:j2099.
- 12 Gorelick MH, Knight S, Alessandrini EA, *et al*. Pediatric Emergency Care Applied Research Network. *Acad Emerg Med* 2007;14:646–52.
- 13 McColm D, Karcz A. Comparing manual and automated coding of physicians quality reporting initiative measures in an ambulatory EHR. *J Med Pract Manage* 2010;26:6–12.
- 14 Sheikhshahi S, Miotto R, Dudley JT, *et al*. Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. *JMIR Med Inform* 2019;7:e12239.
- 15 Roberts K, Chin AT, Loewy K, *et al*. Natural language processing of clinical notes enables early inborn error of immunity risk ascertainment. *J Allergy Clin Immunol Glob* 2024;3:100224.
- 16 Song J, Topaz M, Landau AY, *et al*. Natural Language Processing to Identify Home Health Care Patients at Risk for Becoming Institutionalized With No Evident Advance Directives or Surrogates. *J Am Med Dir Assoc* 2024;25:105019.
- 17 Omar M, Naffaa ME, Glicksberg BS, *et al*. Advancing rheumatology with natural language processing: insights and prospects from a systematic review. *Rheumatol Adv Pract* 2024;8:rkae120.
- 18 Hussein KI, Chan L, Van Vleck T, *et al*. Natural language processing to identify patients with cognitive impairment. *Geriatric Medicine [Preprint]* 2022.
- 19 Hossain E, Rana R, Higgins N, *et al*. Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review. *Comput Biol Med* 2023;155:106649.
- 20 Owen RK, Lyons J, Akbari A, *et al*. Effect on life expectancy of temporal sequence in a multimorbidity cluster of psychosis, diabetes, and congestive heart failure among 1.7 million individuals in Wales with 20-year follow-up: a retrospective cohort study using linked data. *Lancet Public Health* 2023;8:e535–45.