

**Agitation prevalence in people with dementia in Australian residential aged care facilities:
Findings from machine learning of electronic health records**

Authors: Yunshu Zhu¹, Ting Song¹, Zhenyu Zhang¹, Chao Deng^{2,3}, Mohammad Alkhalaf¹,
Wanqing Li⁴, Mengyang Yin⁵, Hui-Chen (Rita) Chang⁶, Ping Yu^{1,2*}

¹Centre for Digital Transformation, School of Computing and Information Technology,
University of Wollongong, Wollongong, New South Wales, Australia

²Illawarra Health and Medical Research Institute

³School of Medicine, University of Wollongong, Wollongong, New South Wales, Australia

⁴Artificial Intelligence Research Centre, School of Computing and Information Technology,
University of Wollongong, Wollongong, New South Wales, Australia

⁵Systems and Reporting Residential Care, Catholic Healthcare Ltd, Macquarie Park, New South
Wales, Australia

⁶School of Nursing, University of Wollongong, Wollongong, New South Wales, Australia

*: Correspondence Author

Acknowledgement

We are grateful for the aged care organization that shared with us the de-identified electronic health records. This provided us with the opportunity to conduct this significant data mining research project.

Suggested Citation: Yunshu Zhu, Ting Song, Zhenyu Zhang, Chao Deng, Mohammad Alkhalaf, Wanqing Li, Mengyang Yin, Hui-Chen (Rita) Chang, Ping Yu (2022) Agitation prevalence in people with dementia in Australian residential aged care facilities: Findings from machine learning of electronic health records. *Journal of Gerontological Nursing*

Agitation prevalence in people with dementia in Australian residential aged care facilities:

Findings from machine learning of electronic health records

Abstract: Using a suite of artificial intelligence technologies, the current study sought to determine the prevalence of agitated behaviors in people with dementia in residential aged care facilities (RACFs) in Australia. Computerised natural language processing allowed extraction of agitation instances from the free-text nursing progress notes, a component of the electronic health records in RACFs. In total, 59 observable agitated behaviors were found. No difference was found in dementia prevalence between female and male clients (44.1%), across metropolitan and regional facilities ($42.1 \pm 17.9\%$), nor for agitation prevalence in dementia ($76.5 \pm 18.4\%$). The top 10 behaviors were resisting, wandering, speaking in excessively loud voice, pacing, restlessness, pushing, shouting, complaining, frustration, and using profane language. Four to 17 agitated behaviors co-existed in 53% of people with dementia agitation, indicating high caregiver burden in these RACFs. Improving workforce training and redesigning care models are urgent for sustainability of dementia care in RACFs.

Keywords: Agitation prevalence, Agitated behavior, Dementia, Nursing home, Residential aged care, Artificial intelligence, Ontology, Knowledge graph, Natural language processing, Machine learning

1 **Agitation prevalence in people with dementia in Australian residential aged care facilities:**

2 **Findings from machine learning of electronic health records Introduction**

3 Dementia is one of the major causes of disability and dependency among older people (World
4 Health Organization, 2021). Worldwide, around 55 million people live with dementia. This
5 number is estimated to increase to 78 million in 2030 and 139 million in 2050. In Australia,
6 dementia was the 2nd leading cause of death in 2019, and the leading cause of death for women,
7 accounting for 9.5% of all deaths (Australian Institute of Health and Welfare (AIHW), 2020). The
8 estimated health expenditure on dementia care in Australia is AU\$15 billion (AIHW, 2021a). The
9 physical, psychological, social and economic impacts of dementia go beyond people with dementia,
10 putting a heavy burden on their carers, families and society at large.

11 Residential aged care facilities (RACFs) provide essential care services for older
12 Australians who lose the ability to live independently in the community, including those with
13 dementia (AIHW, 2021a). In 2019-20, over 244,000 people lived in RACFs, and 54% of these
14 people had dementia. People with dementia often manifest agitated behaviors for various reasons,
15 such as changes taking place in their brain, changes occurring in their environment, their health
16 and medication, their unmet needs, and actions of their carers (Kolanowski et al., 2017). Agitated
17 behavior is one of the most disruptive neuropsychiatric symptoms of people with dementia
18 (Watson and Hatcher, 2021). It encompasses a variety of non-purposeful physical and verbal
19 behaviors such as wandering, repetitive questioning, cursing, screaming, hitting, shouting, spitting,
20 and destroying property (Halpern et al., 2019; Zhang et al., 2020). These behaviors have negative
21 effects on the cognitive performance and quality of life of people with dementia (Livingston et al.,
22 2017; Watson and Hatcher, 2021), cause psychological and physical burdens for caregivers, and
23 increase the cost of care (Hiyoshi-Taniguchi et al., 2018). Therefore, understanding the prevalence

1 of agitated behavior is important in many aspects: it will help aged care staff to better understand
2 people with dementia, their care needs and to provide person-centred care; it will facilitate
3 assessment of severity of dementia and the effectiveness of behavioral management strategies
4 including pharmacological and non-pharmacological (e.g., music therapy) approaches (Hendriks
5 et al., 2014); it will inform researchers and policy-makers about its burden, and thus the effort or
6 workload of managing these behaviors (Gilbert et al., 2017); it will also contribute to staff training
7 and a better staff skill mix for best client outcomes. Most importantly, it will improve the safety
8 and wellbeing of people with dementia (Panca et al., 2019).

9 Due to the high cost of human resources required to collect data or analyze nursing
10 observations recorded in free-text nursing notes, the prevalence of agitated behaviors in dementia
11 has rarely been comprehensively studied. To the best of our knowledge, the most reliable data was
12 collected using staff report rating scales, or direct behavioral observation. Participant self-reporting
13 scores were gathered by the Neuropsychiatric Inventory in a US study in 2008 (Steinberg et al.,
14 2008), which reported the prevalence of agitation in 70%-80% of people with dementia living in
15 RACFs. This study was further quoted in a report of the European Union-North American Clinical
16 Trials in Alzheimer's Disease Task Force for the identification of better outcome measures for
17 improving treatment of agitation in dementia (Sano et al., 2018).

18 Kverno et al. (2008) studied the prevalence of behavioral changes in dementia by reviewing
19 the medical records of 123 people living in RACFs in the USA. They found agitation to be the
20 most frequently observed behavioral change and it occurred in 50.4% of people with dementia.
21 This was followed by depression (45.5%) and withdrawal/lethargy (43.1%). A recent interview
22 study conducted with 11 aged care staff identified factors influencing the management of agitated
23 behavior at two RACFs in Australia (Watson & Hatcher, 2021). Participants reported high

1 frequency of managing this behavior, at least once per shift. It was most frequently manifested as
2 wandering, restlessness, and aggression.

3 Notably, these studies applied manual methods to extract the relevant information and
4 focused on the overall prevalence of agitated behavior and its relationship with other factors (e.g.,
5 dementia severity, treatment types, medications, and other behavioral changes in dementia). To
6 date, there is little reporting of the specific agitated behaviors and their rate of occurrence despite
7 the fact that these behaviors are likely observed and documented by aged care staff and other
8 healthcare providers in human natural language in Electronic Health Records (EHRs) (Halpern et
9 al., 2019; Kverno et al., 2008). These EHRs usually store large longitudinal data that include
10 demographics, progress notes, diagnoses, incident and accident reports, residents' forms and charts,
11 24-hour shift handover reports, assessments, care plans and documents for reimbursement of
12 funding from the Australian Federal Government (Munyisia et al., 2012). Due to the barrier that
13 machines cannot understand the natural language written in a free and unstructured format, the
14 valuable information in EHRs has not been effectively used to generate useful insights to guide
15 aged care practices for people with dementia

16 This barrier has recently been broken down by Natural Language Processing (NLP) - a
17 technique that employs one or more levels of linguistic analysis to translate a natural language
18 description into a machine-processable format (Iroju & Olaleke, 2015). NLP technology has been
19 widely used in the health domain to find and extract specific information from unstructured natural
20 language in EHRs (Juhn & Liu, 2020; Wen et al., 2021). For example, Halpern et al. (2019) used
21 a NLP system to analyze EHRs of 320,886 people with dementia to estimate the prevalence of
22 agitated behavior in these people living in a community in the USA. They reported a 44.6%
23 prevalence of agitated behavior in this population. Agitated behavior was most prevalent among

1 people with moderate-to-severe (74.6%), and severe Alzheimer disease/dementia (68.3%), and
2 least prevalent for mild Alzheimer disease/dementia (56.4%). They reported the common agitated
3 behaviors, including agitation (41.3%), restlessness (23.3%), anger (11.8%), harm to self (6.1%),
4 emotional disturbance (6.3%), and wandering (6.7%).

5 However, few comprehensive studies have been published on the prevalence of specific
6 agitated behaviors in people with dementia living in RACFs. To address this gap, this study aimed
7 to estimate the prevalence of agitated behaviors in people with dementia living in Australian
8 RACFs. We applied NLP technology to extract and identify specific agitated behaviors from the
9 unstructured nursing progress notes in the EHRs.

10 **Materials and Methods**

11 **Ethics Approval**

12 This study had been approved by the Human Research Ethics Committee of the University
13 of Wollongong (Registration number 2019/159).

14 **Study Design**

15 This retrospective cohort study analyzed the nursing progress notes recorded by registered
16 nurses, enrolled nurses, care workers, and other healthcare professionals (e.g., physicians and
17 therapists) in the process of delivering aged care services on a regular basis to identify the
18 prevalence and symptoms of agitation in people with dementia in Australian RACFs.

19 **Data Source**

20 Two datasets were acquired from the EHR database of a large aged care organization
21 running 40 RACFs in New South Wales and Queensland, Australia in 2019. The datasets included
22 more than 890,000 notes of 3,528 de-identified clients, which included their demographic data and
23 nursing progress notes. Demographic data were structured, including sequence number, age,

1 gender, and residing facility sequence number. Nursing progress notes were unstructured, free-text
2 data, including residents' daily activities, clinical observations, assessment of care needs
3 (including risk factors) and nursing interventions.

4 **Data Processing and Analysis**

5 Data processing included three distinct phases: (1) data pre-processing, (2) rule-based NLP
6 algorithm development and evaluation, and (3) data analysis conducted using three open-source
7 libraries – Natural language toolkit (NLTK) (version 3.5) (Bird et al., 2009), Pandas (version 1.2)
8 (McKinney & computing, 2011), and ScispaCy (version 0.3) (Neumann et al., 2019) in Python
9 programing language (version 3.7).

10 ***Phase 1. Data pre-processing***

11 The free-text data contain a substantial amount of (1) irrelevant or duplicate notes, (2) non-
12 alphanumeric characters and punctuation, and (3) stop words, such as “a”, “the”, “is”, and “are” .
13 To remove this noise, the NLTK tokenizer was used for data cleaning. The cleaned data were
14 transferred into CSV format and saved in Microsoft Excel spreadsheet (Elliott et al., 2006).

15 ***Phase 2. Rule-based NLP algorithm development and evaluation***

16 To conduct a comprehensive search in the dataset, two rule-based algorithms were created,
17 one to code dementia, and the other to code agitation related keywords, respectively. A preliminary
18 list of terms for dementia was determined using a large, standardized database of health
19 terminology, Unified Medical Language System (Bodenreider, 2004). The diagnosis notes were
20 also referenced to detect dementia-related terms. After independent review and verification by two
21 domain experts, the algorithm for identifying clients with dementia was finally determined.

22 A machine-understandable ontology, DRANPTO (Zhang et al. 2020), was applied to
23 identify agitation-related terms. The DRANPTO is the first comprehensive knowledge

1 representation of non-pharmacological management for agitation in dementia, developed and
2 evaluated by domain experts (Zhang et al., 2020). It summarizes the observable agitated behaviors,
3 and classifies them into three categories: physical behavior (34 observable behaviors), verbal
4 behavior (27 observable behaviors) and emotional distress (6 observable behaviors). The terms
5 included all the agitation concepts and their synonyms in the ontology (see Supplemental Material
6 1).

7 A gold standard was developed by three researchers to identify the diverse language
8 patterns used to describe agitated behavior. First, 30 random nursing progress notes were extracted
9 and individually analyzed by three researchers. The inter-rater reliability was tested using IBM
10 SPSS Statistics (version 28) (Field, 2013). Afterwards, 300 nursing progress notes were manually
11 annotated and the agitated behaviors were extracted to develop a rule-based algorithm. The
12 algorithm was applied in computer language Python to iteratively detect each agitated behavior
13 using the original words, definitions, and synonyms in DRANPTO, as well as the relevant
14 expressions annotated by the three researchers. Each agitated behavior has one to thirteen rules
15 (see Supplemental Material 2 for a representative sample of these rules).

16 The two rule-based algorithms were used to search the whole nursing notes. An agitated
17 behavior would be extracted when at least one of the rules was detected. Finally, the performance
18 of the rule-based NLP algorithms was evaluated on 300 annotated nursing notes by both NLP
19 algorithm and human to assess sensitivity, specificity and F-Score of the NLP algorithm. The
20 evaluation of NLP algorithm used the results of the true positives (TP, instances of agitated
21 behavior detected by both the NLP algorithm and the manual annotators), false positives (FP,
22 instances of agitated behavior detected by the algorithm but not the manual annotators), true
23 negatives (TN, instances that are not detected as agitated behavior by both the algorithm and the

1 manual annotators), and false negatives (FN, instances of agitated behavior that were not detected
2 by the algorithm but were detected by the manual annotators) (Midgley et al., 1969):

3
$$\text{Sensitivity} = \text{TP}/(\text{TP}+\text{FN}) \times 100\%$$

4
$$\text{Specificity} = \text{TN}/(\text{TN}+\text{FP}) \times 100\%$$

5
$$\text{F-Score} = \text{TP}/(\text{TP} + 0.5(\text{FP} + \text{FN}))$$

6 ***Phase 3. Data analysis***

7 Descriptive statistics were conducted using IBM SPSS Statistics (version 28) (Field, 2013),
8 which included characteristics of people in the RACFs, comparisons of the prevalence of dementia,
9 and agitated behavior in dementia in different gender and age groups, manifestation of agitated
10 behaviors, and the top 25 agitated behaviors.

11 **Results**

12 **The Accuracy of the Rule-based Natural Language Processing Model**

13 Our rule-based natural language processing model achieved 89% of sensitivity, meaning
14 that 89% of people with agitated behavior were correctly identified. Its specificity is 68%, meaning
15 the algorithm's chance of accurately differentiating people without agitated behavior is 68%. F-
16 score of 0.89 suggested an 89% of chance of achieving perfect precision and recall.

17 **The Prevalence of Dementia and Agitated Behavior in People with Dementia in**

18 **RACFs**

19 In total there were 3,528 clients in 40 RACFs, including 2,267 (64%) female and 1,261
20 (36%) male clients. Twenty-six RACFs were in the metropolitan areas, including Sydney and
21 Brisbane; 14 RACFs were located in the regional cities, such as Wollongong, Newcastle, and
22 Grafton. Dementia prevalence was similar between the female (44.2%, 1,003/2,267) and the male
23 clients (43.9%, 553/1261), as well as the prevalence of agitated behavior in dementia (female:

1 78.6%, 788/1,003; male: 79.6%, 440/553). There was a similar number of clients (45±29) in each
 2 RACF in the metropolitan and regional areas. Again, no significant difference was found in the
 3 prevalence of dementia (42.1±17.9%), nor the prevalence of agitated behavior in people with
 4 dementia in the metropolitan and regional RACFs (76.5±18.4%).

5 **The Prevalence of Dementia, and Agitated Behavior in People with Dementia in**

6 **Different Gender and Age Groups**

7 Both female and male clients in the same age group showed a similar trend of dementia
 8 occurrence, which peaked at above 45% in the 76 to 95 years age group (see Table 1). A similar
 9 trend was observed for the agitated behaviors, which were commonly observed at about 72% of
 10 people with dementia in the under 65 years age group; then increased to above 80% in the 66 to
 11 85 age group in both female and male people with dementia; afterwards it dropped slightly to 70-
 12 80% in the over 85 years age group.

13 **Table 1**

14 *The number and percentage of clients with dementia, and agitated behavior in dementia in*
 15 *different gender and age groups*

Gender	Age (years)	No. of Clients	No. with Dementia (%)	No. with Agitated Behavior in Dementia (%)
Female	40 - 65	58	18 (31.0%)	13 (72.2%)
	66 - 75	207	63 (30.4%)	52 (82.5%)
	76 - 85	570	277 (48.6%)	225 (81.2%)
	86 - 95	1143	535 (46.8%)	420 (78.5%)
	95+	289	110 (38.1%)	78 (70.9%)
	Sub-total	2267	1003 (44.2%)	788 (78.6%)
Male	44 - 65	50	11 (22%)	8 (72.7%)
	66 - 75	207	82 (39.6%)	66 (80.5%)
	76 - 85	440	199 (45.2%)	165 (82.9%)
	86 - 95	502	239 (47.6%)	184 (77.0%)
	95+	62	22 (35.4%)	17 (77.3%)
	Sub-total	1261	553 (43.9%)	440 (79.6%)

Total	3528	1556 (44.1%)	1228 (78.9%)
--------------	-------------	---------------------	---------------------

1
2
3
4
5
6
7
8
9
10
16
17
18
19
20
21

The Manifestation of Agitated Behavior in People with Dementia

In total, 59 observable agitated behaviors were identified (see Supplemental Material 3). The most manifested agitated behavior was physical agitation (64.5%, see Table 2), followed by verbal disruption (38.5%), and emotional distress (9.3%). Ten commonly observed agitated behaviors were resisting, wandering, speaking in excessively loud voice, pacing, restlessness, pushing, shouting, complaining, frustration, and using profane language (see Figure 1). There was no gender difference in the occurrence of any specific behavior (see Figure 2).

Table 2

The manifested agitated behavior in male and female clients with dementia

Gender	No. with Agitated Behavior	No. (%) with Physical Agitation	No. (%) with Verbal Disruption	No (%) with Emotional Distress
Female	788	509 (64.6%)	303 (38.5%)	70 (8.8%)
Male	440	283 (64.3%)	170 (38.6%)	44 (10.0%)
Total	1228	792 (64.5%)	473 (38.5%)	114 (9.3%)

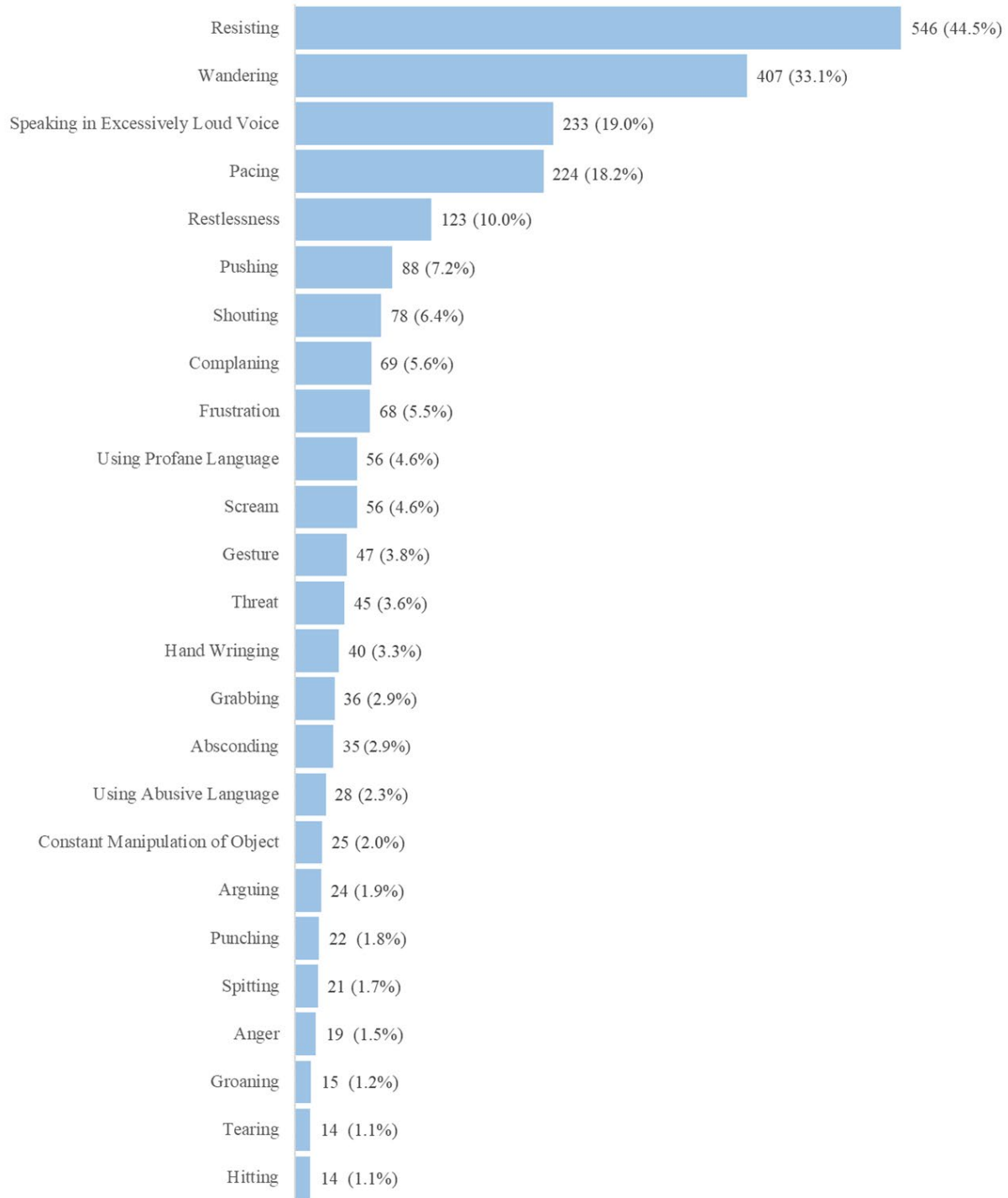
The number of manifested agitated behaviors in one person

One to 17 agitated behaviors were found in one client living with dementia, irrespective of gender (see Figure 3). One agitated behavior was found in 25% of these people, two behaviors in 13%, three behaviors in 9%, four behaviors in 6%, five to nine behaviors in 21%, and ten to 17 behaviors in 26%.

Figure 1

1 *Top 25 agitated behaviors and their percentages in all clients with agitated behaviors in*

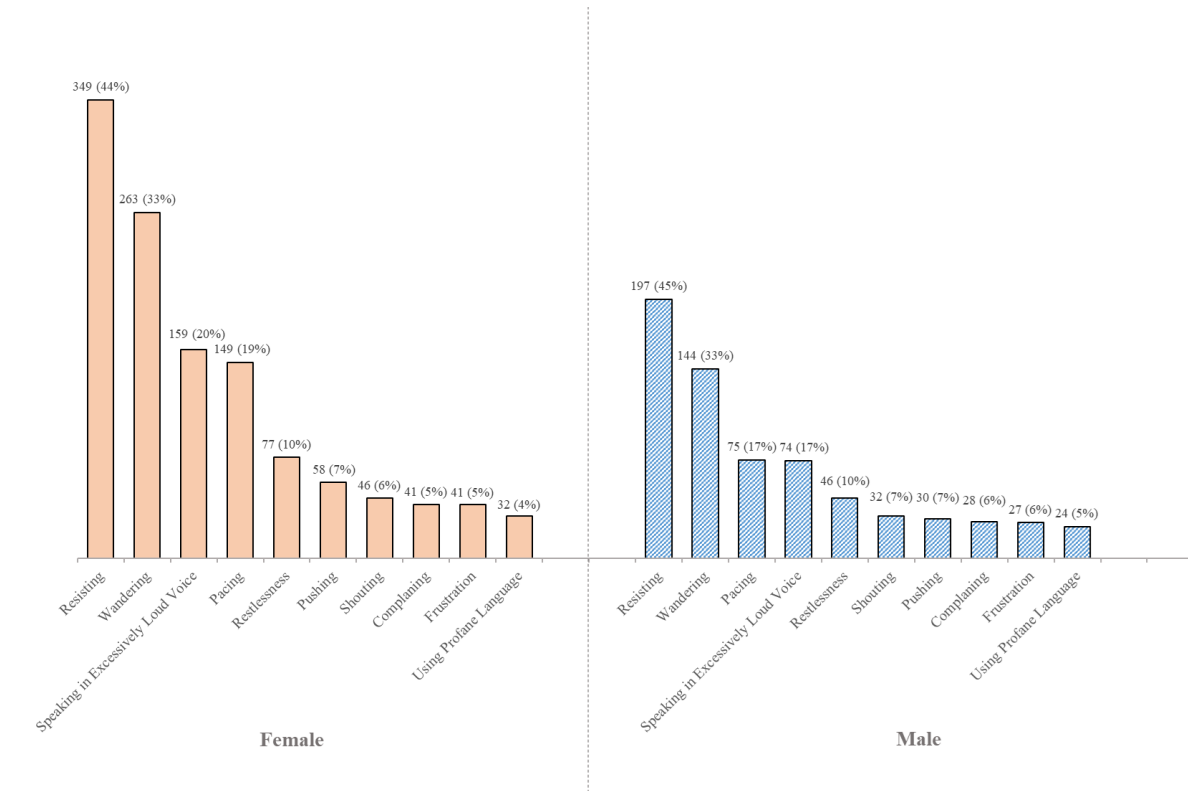
2



3

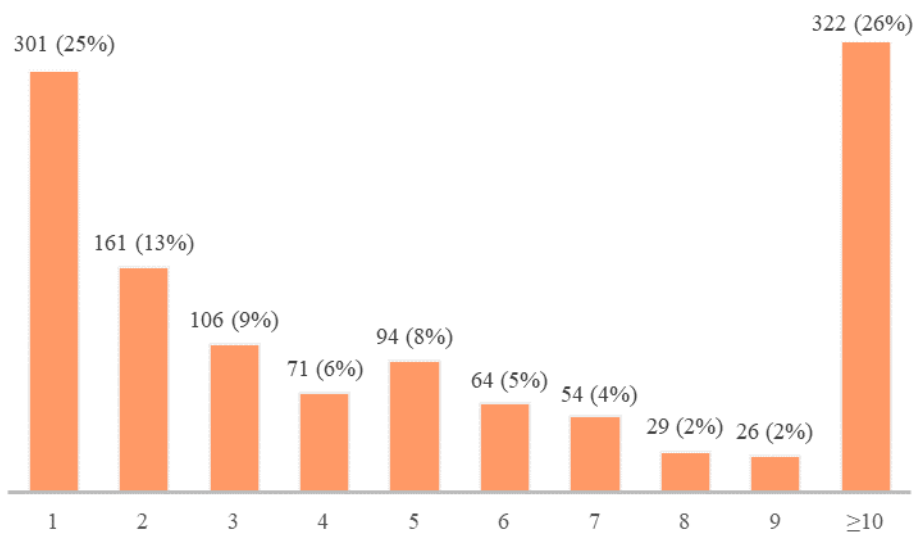
4 **Figure 2**

1 Comparison of the number of female or male clients with top 10 agitated behaviors (% of clients with the
 2 behavior / number of female or male people with agitated behavior in dementia)



4 **Figure 3**

5 The number (%) of agitated behaviors in all people with agitated behavior in dementia



1 **Discussion**

2 Despite the perceived high prevalence, agitated behavior is less frequently studied in
3 nursing home residents with dementia (Hendriks et al., 2014). Lack of clear knowledge about the
4 specific manifestation of agitated behavior of individual clients and their causative factors has
5 jeopardized the provision of individualized, person-centred nursing care (Watson & Hatcher, 2021).
6 Therefore, this study makes two significant contributions to gerontology and gerontological
7 nursing: (1) contributing novel machine learning and NLP methods to uncover the key health care
8 information from the rich nursing progress notes; and (2) contributing knowledge about the
9 prevalence of specific agitated behaviors in people with dementia in RACFs.

10 **Contributing Novel Machine Learning and NLP Methods to Uncover the Key**
11 **Health and Care Information from the Rich Nursing Progress Notes**

12 Utilization of the free-text nursing records in an EHR database to access nursing
13 observations at a large sample size was a strength of this study. Halpern et al. (2019) conducted
14 the first study that applied NLP technology to process large amount of EHR data to estimate the
15 prevalence of agitated behavior in people with dementia living in the community . To the best of
16 our knowledge, our study is the first to apply a suite of artificial intelligence (AI) technologies, i.e.,
17 ontology, rule-based algorithm development and validation, and computerized natural language
18 processing to uncover the key health and care knowledge from the routinely documented, free-text
19 nursing progress notes. It then applied descriptive statistics to generate the evidence-based findings
20 about prevalence of agitated behaviors in people with dementia in residential aged care facilities.

21 First, it used the standardized terminology in the established ontology – DRANPTO - to
22 represent agitated behaviors. This allowed mapping of the various free-text descriptions in the
23 routinely captured nursing progress notes to these standardized terms. Second, we developed rules

1 for automating the human decision-making process of detecting each manifestation of agitated
2 behaviors from the diverse nursing descriptions. Third, we applied these rules to automatically
3 extract the corresponding terms matching the agitated behaviors from the nursing progress notes.
4 Finally, we followed the common machine learning evaluation criteria, sensitivity, specificity, and
5 F-score, to validate the performance of these rule-based NLP algorithms. The above data mining
6 methods and workflow can be replicated to extract knowledge from the free-text data.

7 **Contributing knowledge about the prevalence of specific agitated behaviors in** 8 **people with dementia in RACFs**

9 Many studies on agitated behavior of people with dementia collect evidence based on
10 behavioral observation or questionnaire survey, which is either labor intensive or subject to self-
11 reporting bias. For example, Kverno et al. (2008) manually reviewed the medical records of people
12 with dementia in nursing homes, but did not distinguish the specific agitated behaviors. For the
13 first time, this study generated a detailed, comprehensive list of the specific agitated behaviors for
14 people with dementia in RACFs.

15 The validity of our finding is supported by the Australian national data, reporting two in
16 three people using residential aged care were women (AIHW, 2021b). The agitation prevalence
17 rate of 79% found in this study is in agreement with earlier study results; 70-80% of people with
18 dementia in RAC (Steinberg et al., 2008), and slightly higher than 76% found in a cross-sectional
19 study of people with dementia in the community (Jones et al., 2021). As expected, this rate is
20 much higher than the prevalence rate of 44.6% in people with dementia living in the community
21 (Halpern et al., 2019). For the first time, we found similar manifestations and ratio of agitated
22 behaviors in both female and male people with dementia in RACFs: i.e., about 45% of people with
23 dementia had the resisting behavior, followed by 33% wandering, speaking in excessively loud

1 voice (female: 20%; male: 17%), and pacing (17%) in male clients. Watson and Hatcher (2021)
2 found the most frequently manifested behavior were wandering, restlessness, and aggression.
3 Therefore, both studies found the common behavior of wandering. Our concept of resisting
4 appeared to be similar with the term aggression used by Watson and Hatcher (2021); however, this
5 cannot be fully decided as their study did not provide definition of terms.

6 A previous study found that four or more behavioral and psychological symptoms of
7 dementia have strong predictive value for caregiver burden and depression (Arthur et al., 2018).
8 In particular, agitation in dementia is associated with increased health care resource use and cost
9 (Jones et al., 2021). In this study, four to 17 behavioral changes were found in 53% of people with
10 agitation in dementia. This may suggest the high risk of caregiver burden and depression in these
11 residential aged care facilities. This evidence suggests that increasing amount of care staff, staff
12 training and appropriate skill mix are necessary for improving dementia care and sustainability of
13 aged care services in Australia.

14 This study demonstrates the power of AI technology in processing large amount of
15 routinely collected nursing observational data stored in EHR to generate key health and care
16 information. It sets up a new direction for AI-driven, evidence-based dementia research to identify
17 the determinants of specific BPSD, the progression of dementia and BPSD presentation, the
18 frequency of occurrence and effectiveness / or lack of it of various nursing care options.

19 The method developed in this study can be applied in various fields of gerontological
20 nursing to improve understanding of older people's health and wellbeing, the corresponding
21 nursing actions and outcomes in various healthcare setting. Once the digital platform and process
22 is well-established, it will open new pathway for answering many research questions in a relatively
23 short time frame without the need to invest large human and time resources in the traditional

1 research approach, e.g., randomized controlled trial. In the future, we will apply similar methods
2 to identify the relevant nursing actions to handle the identified agitated behaviors in dementia and
3 the outcomes.

4 This study has several limitations. First, as in any secondary study, accuracy of the findings
5 relies on the level of accuracy and completeness of the original EHR (Halpern et al., 2019), which
6 we could not ascertain. As found by Halpern et al. (2019), a symptom could be recorded in diverse
7 writing styles, causing a challenge for the NLP algorithm to capture all variations. We carefully
8 conducted manual coding by three researchers to capture all the variations. Second, the less
9 satisfactory specificity (68%), the fair performance of algorithms indicated that certain clients
10 without agitation behavior may be classified otherwise, likely because the algorithm did not
11 differentiate a question that may ask ‘Whether a person has agitation?’ thus causing a false positive
12 answer. Otherwise, a satisfactory level of sensitivity (89%) and F-score (0.89) indicated that 89%
13 of individuals with agitation were identified, and the algorithm achieved precision and recall in
14 89% of findings. We will continue to improve machine learning performance.

15 **Conclusion**

16 Applying the artificial intelligence methods, for the first time, this study contributes to a
17 comprehensive evaluation of the prevalence of agitation and specific agitated behaviors in people
18 with dementia in 40 Australian RACFs. The study finds a similar level of prevalence of dementia
19 and agitated behavior in both female and male clients, and across the metropolitan and regional
20 facilities. The top 10 symptoms were resisting, wandering, speaking in excessively loud voice,
21 pacing, restlessness, pushing, shouting, complaining, frustration, and using profane language. As
22 53% of people with agitation in dementia manifested 4 to 17 types of behavioral changes, it
23 appears that caregiver burden was high in these RACFs. The research methods and knowledge

1 generated in this study is useful for the future design of data-driven practice improvement in
2 aged care in particular, and nursing in general.

3 **Author Contributions**

4 YZ contributed to study conception, data mining, statistical analysis and manuscript
5 drafting. YZ, TS, ZZ and MY contributed to manual coding of nursing notes, algorithm
6 validation and manuscript drafting. CD, WL and HC contributed to manuscript drafting. MA
7 contributed to the design of the data mining method. PY contributed to research conception,
8 guidance on data analytics, algorithm validation and manuscript drafting. All of the authors
9 approved the manuscript.

10 **Acknowledgement**

11 We are grateful for the aged care organization that shared with us the de-identified electronic
12 health records, and provided us with the opportunity to conduct this significant data mining
13 research project.

References

- 1
2 Arthur, P. B., Gitlin, L. N., Kairalla, J. A., & Mann, W. C. (2018). Relationship between the
3 number of behavioral symptoms in dementia and caregiver distress: what is the tipping
4 point? *Int Psychogeriatr*, 30(8), 1099-1107. <https://doi.org/10.1017/s104161021700237x>
- 5 Australian Institute of Health and Welfare. (2020). *Dementia: Data*.
6 <https://www.aihw.gov.au/reports-data/health-conditions-disability-deaths/dementia/data>
- 7 Australian Institute of Health and Welfare. (2021a). *Dementia in Australia: Residential aged*
8 *care*. [https://www.aihw.gov.au/reports/dementia/dementia-in-aus/contents/aged-care-and-](https://www.aihw.gov.au/reports/dementia/dementia-in-aus/contents/aged-care-and-support-services-used-by-people-with-dementia/residential-aged-care)
9 [support-services-used-by-people-with-dementia/residential-aged-care](https://www.aihw.gov.au/reports/dementia/dementia-in-aus/contents/aged-care-and-support-services-used-by-people-with-dementia/residential-aged-care)
- 10 Australian Institute of Health and Welfare. (2021b). *GEN Aged Care Data: People using aged*
11 *care*. <https://www.gen-agedcaredata.gov.au//Topics/People-using-aged-care>
- 12 Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text
13 with the natural language toolkit. *O'Reilly Media, Inc*.
14 <https://doi.org/10.2310/6650.2006.05038>
- 15 Bodenreider, O. (2004). The unified medical language system (UMLS): integrating biomedical
16 terminology. 32(suppl_1), D267-D270. <https://doi.org/10.1093/nar/gkh061>
- 17 Elliott, A. C., Hynan, L. S., Reisch, J. S., & Smith, J. P. (2006). Preparing data for analysis using
18 Microsoft Excel. *Journal of investigative medicine*, 54(6), 334-341.
19 <https://doi.org/10.2310/6650.2006.05038>
- 20 Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. *sage*.
- 21 Gilbert, C., Wilcock, J., Thuné-Boyle, I., & Iliffe, S. (2017). A comparison of service use by
22 people with dementia in two samples a decade apart. *Dementia*, 16(1), 96-107
23 <https://doi.org/10.1177/1471301215581504>

- 1 Halpern, R., Seare, J., Tong, J., Hartry, A., Olaoye, A., & Aigbogun, M. S. (2019). Using
2 electronic health records to estimate the prevalence of agitation in Alzheimer
3 disease/dementia. *International journal of geriatric psychiatry*, 34(3), 420-431.
4 <https://doi.org/10.1002/gps.5030>
- 5 Hendriks, S. A., Smalbrugge, M., Hertogh, C. M., & van der Steen, J. T. (2014). Dying with
6 dementia: symptoms, treatment, and quality of life in the last week of life. *Journal of*
7 *pain and symptom management*, 47(4), 710-720.
8 <https://doi.org/10.1016/j.jpainsymman.2013.05.015>
- 9 Hiyoshi-Taniguchi, K., Becker, C. B., & Kinoshita, A. (2018). What behavioral and
10 psychological symptoms of dementia affect caregiver burnout? *Clinical gerontologist*,
11 41(3), 249-254. <https://doi.org/10.1080/07317115.2017.1398797>
- 12 Iroju, O. G., & Olaleke, J. O. (2015). A systematic review of natural language processing in
13 healthcare. *International Journal of Information Technology and Computer Science*, 7(8),
14 44-50. <https://doi.10.5815/ijitcs.2015.08.07>
- 15 Jones, E., Aigbogun, M. S., Pike, J., Berry, M., Houle, C. R., & Husbands, J. (2021). Agitation in
16 dementia: Real-world impact and burden on patients and the healthcare system. *J*
17 *Alzheimers Dis*, 83(1), 89-101. <https://doi.org/10.3233/jad-210105>
- 18 Juhn, Y., & Liu, H. (2020). Artificial intelligence approaches using natural language processing
19 to advance EHR-based clinical research. *Journal of Allergy and Clinical*
20 *Immunology*, 145(2), 463-469. <https://doi.org/10.1016/j.jaci.2019.12.897>
- 21 Kolanowski, A., Boltz, M., Galik, E., Gitlin, L. N., Kales, H. C., Resnick, B., ... & Scerpella, D.
22 (2017). Determinants of behavioral and psychological symptoms of dementia: A scoping

1 review of the evidence. *Nursing Outlook*, 65(5), 515-529.
2 <https://doi:10.1016/j.outlook.2017.06.006>

3 Kverno, K. S., Rabins, P. V., Blass, D. M., Hicks, K. L., & Black, B. S. (2008). Prevalence and
4 treatment of neuropsychiatric symptoms in advanced dementia. *Journal of*
5 *Gerontological Nursing*, 34(12), 8-14.

6 Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., &
7 Mukadam, N. (2017). Dementia prevention, intervention, and care. *The Lancet*,
8 390(10113), 2673-2734.

9 McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics.
10 *Python for high performance and scientific computing*, 14(9), 1-9.

11 McLean, A. H., Leibing, A., & Cohen, L. (2006). 7. Coherence without Facticity in Dementia:
12 The Case of Mrs. Fine. *In Thinking about dementia* (pp. 157-179). Rutgers University
13 Press.

14 Midgley, A. R., Niswender, G. D., & Rebar, R. W. (1969). Principles for the assessment of the
15 reliability of radioimmunoassay methods (precision, accuracy, sensitivity, specificity).
16 *European Journal of Endocrinology*, 62(1_Suppl), S163-S184.

17 Munyisia, E. N., Yu, P., & Hailey, D. (2012). The impact of an electronic nursing documentation
18 system on efficiency of documentation by caregivers in a residential aged care facility.
19 *Journal of clinical nursing*, 21(19pt20), 2940-2948. [https://doi.10.1111/j.1365-](https://doi.10.1111/j.1365-2702.2012.04157.x)
20 [2702.2012.04157.x](https://doi.10.1111/j.1365-2702.2012.04157.x)

21 Neumann, M., King, D., Beltagy, I., & Ammar, W. (2019). Scispacy: Fast and robust models for
22 biomedical natural language processing. *arXiv preprint arXiv:1902.07669*.

1 Panca, M., Livingston, G., Barber, J., Cooper, C., La Frenais, F., Marston, L., & Hunter, R. M.
2 (2019). Healthcare resource utilisation and costs of agitation in people with dementia
3 living in care homes in England-the managing agitation and raising quality of life in
4 dementia (MARQUE) study. *PloS one*, *14*(2), e0211953.
5 <https://doi.org/10.1371/journal.pone.0211953>

6 Sano, M., Soto, M., Carrillo, M., Cummings, J., Hendrix, S., Mintzer, J., Porsteinsson, A.,
7 Rosenberg, P., Schneider, L., Touchon, J., Aisen, P., Vellas, B., & Lyketsos, C. (2018).
8 Identifying Better Outcome Measures to Improve Treatment of Agitation in Dementia: A
9 Report from the EU/US/CTAD Task Force. *J Prev Alzheimers Dis*, *5*(2), 98-102.
10 <https://doi.org/10.14283/jjpad.2018.15>

11 Steinberg, M., Shao, H., Zandi, P., Lyketsos, C. G., Welsh-Bohmer, K. A., Norton, M. C.,
12 Breitner, J. C., Steffens, D. C., & Tschanz, J. T. (2008). Point and 5-year period
13 prevalence of neuropsychiatric symptoms in dementia: the Cache County Study. *Int J*
14 *Geriatr Psychiatry*, *23*(2), 170-177. <https://doi.org/10.1002/gps.1858>

15 Watson, K., & Hatcher, D. (2021). Factors influencing management of agitation in aged care
16 facilities: A qualitative study of staff perceptions. *Journal of Clinical Nursing*, *30*(1-2),
17 136-144. <https://doi.org/10.1111/jocn.15530>

18 Wen, A., Rasmussen, L. V., Stone, D., Liu, S., Kiefer, R., Adekkanattu, P., & Jiang, G. (2021).
19 CQL4NLP: development and integration of FHIR NLP extensions in clinical quality
20 language for EHR-driven phenotyping. In *AMIA Annual Symposium Proceedings* (Vol.
21 2021, p. 624). American Medical Informatics Association.

22 World Health Organization. (2020a). *Dementia: Key facts*. World Health Organization.
23 <https://www.who.int/news-room/fact-sheets/detail/dementia>

1 Zhang, Z., Yu, P., Chang, H. C., Lau, S. K., Tao, C., Wang, N., & Deng, C. (2020). Developing
2 an ontology for representing the domain knowledge specific to non-pharmacological
3 treatment for agitation in dementia. *Alzheimer's & Dementia: Translational Research &*
4 *Clinical Interventions*, 6(1), e12061. <https://doi.org/10.1002/trc2.12061>

5

Supplemental Material 1

1

2 **Agitation terms and definitions**

3 Table 1. Categories, terms and definitions of agitated behaviors in the Dementia-Related

4 Agitation Non-Pharmacological Treatment Ontology (DRANPTO) (Zhang et al. 2020)

Category	Terms	Definition
Physically aggressive behavior	Biting	Using teeth to bite someone aggressively
	Destroying property	Physically damaging or ruining things
	Fighting	Taking part in violent actions involving the exchange of physical blows or the use of tools as weapons
	Grabbing	Grasping someone suddenly and roughly
	Hitting	Using one's hand or an object to strike others
	Inappropriate sexual behavior	Performing a sexual act which is inappropriate within the context of occurrence
	Kicking	Striking someone with the foot
	Punching	Striking someone with the fist
	Pushing	Moving forward by using force to pass people or cause them to move aside
	Resisting	Refusal to be helped with care
	Scratching	Hurting someone's body with one's fingernails
	Self-harm behavior	Hurting or harming oneself without the motive of suicide
	Shoving	Roughly pushing someone with great power
	Slamming	Shutting an object (a door, window, or lid) forcibly and noisily
	Spitting	Ejecting saliva forcibly from one's mouth and inappropriately spitting out, such as spitting at other people or onto the floor
Staring	Looking fixedly at someone or something with one's eyes wide open	
Tearing	Pulling something apart or to pieces with force, such as tearing books	
Throwing Object	Propelling an object with force through the air by a forward motion of the hand and arm	
Physically nonaggressive behavior	Absconding	Behavior of a resident leaving secretly from the long-term care facility in which he/she lives

Constant manipulation of object	Constantly treating with or playing an object with hands
Fidgeting	Making small movements excessively, especially of the hands and feet
Gesture	Making an obscene gesture (a movement or position of the body, especially a hand, the face or the arm) that is considered offensive or vulgar
Hand wringing	Constant twisting and rubbing one's hands
Inappropriate dressing	Wearing clothes that are not suitable or proper in the circumstances
Inappropriate handling object	Treating or managing object(s) in a manner that is not proper in the circumstances, e.g., twisting phone cord, dumping water jug, pulling all tissues out of box
Inappropriate undressing	Taking off clothes in the inappropriate circumstances
Pacing	Walking with regular slow or fast paces, especially without a particular destination
Pointing finger	Behavior of pointing a finger at someone, to blame or accuse one of doing something wrong
Repetitive physical mannerism	Repetitively behaving certain physical actions
Restlessness	Continuously moving characterized by not being able to remain at rest
Rocking	Monotonous rhythmic movements of body from side to side
Rummaging	Searching for lost objects, often in an aimless, repetitive, or fruitless manner
Searching	Behavior of thoroughly scrutinizing, especially in a disconcerting way
Wandering	Meandering, aimless, or repetitive locomotion that exposes an individual to harm

1

2

- 1 Table 2. Agitation verbal behavior terms and definitions from the Dementia-Related Agitation
- 2 Non-Pharmacological Treatment Ontology (DRANPTO) (Zhang et al. 2020)

Category	Terms	Definition
Verbally aggressive behavior	Arguing	Behavior of expressing diverging or opposite views, typically in a heated or angry way
	Complaining	Behavior of expressing feelings of dissatisfaction or annoyance about something
	Cursing	Swearing, especially as an expression of anger or annoyance
	Threat	An expression of an intention to injury, damage, or perform other hostile actions on someone
	Using abusive language	Using extremely rude and insulting language
	Using accusatory language	Using language indicating or suggesting that one believes a person has done something wrong
	Using hostile language	Using unfriendly language
	Using obscene language	Using the language that is morally offensive in a sexual way
	Using profane language	Using the language that is deeply offensive, especially for someone's religious beliefs
Verbally nonaggressive behavior	Ceaseless talking	Constant and unending talking
	Constant repetition of word	Continuously repeating words or sentences
	Constant unwarranted requests for attention	Constant and not justified requests for attention
	Constant unwarranted requests for help	Constant and not justified requests for help
	Constant unwarranted requests for reassurance	Constant and not justified requests for reassurance
	Disorganized speech	Speaking incoherently, e.g., respond to questions with unrelated answers, say illogical things, and shift topics frequently
	Echolalia	Repetition of words or phrases spoken by others
	Groaning	Voicing a deep, inarticulate sound
Grunting	Voicing a low, short guttural sound	

	Howling	Voicing a long, doleful cry or wailing sound
	Making bizarre noise	Making very strange or unusual noise
	Rambling speech	Talking at length in a confused or inconsequential way
	Repetitive Questioning	Continuously and repeatedly asking same questions
	Roaring	Voicing a very loud, deep, prolonged sound
	Screaming	Uttering a long, loud, piercing sound or cry
	Self-talk	Vocalized speech that is directed to oneself or an imaginary recipient
	Shouting	Uttering a sudden loud cry
	Speaking in excessively loud voice	Speaking in a voice which is excessively loud

1
2 Table 3. Agitation emotional distress terms and definitions from the Dementia-Related Agitation
3 Non-Pharmacological Treatment Ontology (DRANPTO) (Zhang et al. 2020)

Category	Terms	Definition
Emotional distress	Anger	A strong emotional feeling of displeasure or hostility
	Frustration	The feeling or state of being frustrated, resulting from being blocked, thwarted, disappointed, or defeated
	Irritability	Abnormal or excessive excitability with easily triggered anger, annoyance, or impatience
	Mood swing	Displaying unpredictable and rapidly changing emotions or moods
	Negativism	State of mind or behavior characterized by extreme skepticism and persistent opposition or resistance to outside suggestions or advice
	Outburst	A sudden, violent expression of emotion

1

Supplemental Material 2

2 **Table 1. A representative sample of the rule-based natural language processing algorithm**

Keywords	Rules	Examples of extraction results
<i>Resisting</i>	“resist* (resisting, resists, resistive, resistance)”, “refus* (refuse, refuses, refusal)”, “decline assistance”, “not consistently following commands” and “is reluctant to”	She refuses staff assistance
<i>Shouting</i>	“shout*(shout, shouts, shouting) out”, “yell*(yells, yelling, yelled)”, “call* (calls, calling) out”	He displays verbal disruption by calling out at staff
<i>Cursing</i>	“curs*(cursing)”, “swearing (at) staff”, “verbal abuse”, “verbal swearing”	He swears at staff

3

Supplemental Material 3

1

2 **Table 1. A complete list of agitated behaviors, number of clients with the symptom, and the**

3 **number (%) of female or male clients with dementia having the symptom.**

Agitated behaviors	No. of Client	No. of Female Client with Dementia (%)	No. of Male Client with Dementia (%)
Resisting	546	349 (44%)	197 (45%)
Wandering	407	263 (33%)	144 (33%)
Speaking in excessively loud voice	233	160 (20%)	73 (17%)
Pacing	224	149 (19%)	75 (17%)
Restlessness	123	78 (10%)	45 (10%)
Pushing	88	58 (7%)	30 (7%)
Shouting	78	47 (6%)	31 (7%)
Complaining	69	41 (5%)	28 (6%)
Frustration	68	40 (5%)	28 (6%)
Using profane language	56	32 (4%)	24 (5%)
Scream	56	44 (6%)	12 (3%)
Gesture	47	33 (4%)	14 (3%)
Threat	45	20 (3%)	25 (6%)
Hand wringing	40	27 (3%)	13 (3%)
Grabbing	36	19 (2%)	17 (4%)
Absconding	35	21 (3%)	14 (3%)
Using abusive language	28	18 (2%)	10 (2%)
Constant manipulation of object	25	11 (1%)	14 (3%)
Arguing	24	19 (2%)	5 (1%)
Punching	22	8 (1%)	14 (3%)
Spitting	21	14 (2%)	7 (2%)
Anger	19	12 (2%)	7 (2%)
Groaning	15	10 (1%)	5 (1%)
Tearing	14	13 (2%)	1
Hitting	14	11 (1%)	3 (1%)
Fighting	13	6 (1%)	7 (2%)
Scratching	13	9 (1%)	4 (1%)
Throwing object	13	8 (1%)	5 (1%)
Rocking	13	11 (1%)	2
Kicking	11	8 (1%)	3 (1%)
Irritability	10	8 (1%)	2
Destroying property	8	4 (1%)	4 (1%)

Self harm behavior	8	4 (1%)	4 (1%)
Inappropriate dressing	8	6 (1%)	2
Requests for reassurance	6	6 (1%)	0
Outburst	6	3	3 (1%)
Cursing	5	2	3 (1%)
Make noise	4	2	2
Repetitive questioning	4	1	3 (1%)
Searching	4	2	2
Using hostile language	3	2	1
Repetition of word	3	2	1
Inappropriate sexual behavior	3	1	2
Mood swing	2	2	0
Biting	2	1	1
Slamming	2	1	1
Fidgeting	2	2	0
Pointing finger	2	1	1
Rummaging	2	1	1
Requests for help	1	1	0
Rambling speech	1	0	1
Negativism	1	1	0
Staring	1	0	1
Inappropriately handling Object	1	0	1
Repetitive physical	1	0	1
Other verbal aggressive behavior	266	170 (22%)	96 (22%)
Other physical aggressive behavior	250	166 (21%)	84 (19%)
Other verbally disruptive behavior	68	47 (6%)	21 (5%)
Other nonaggressive behavior	35	21 (3%)	14 (3%)