

## Tracking housing instability using community supervision case notes: A text mining approach

Chee Seng Chong, Anindita Sarker & Mark Howard

### Aims

This study documents the development and validation of a text mining tool to identify housing-related needs from community supervision case notes.

### Methods

A rule-based text mining tool was developed to identify housing-related keywords and group them into themes. Two indices were developed to summarise patterns of keyword activity. The Point-in-Time (PIT) index provided a discrete snapshot of accommodation needs within a single case note, and the Weighted Moving Average (WMA) index captured trends in those needs over time. Validation involved manual review of keyword capture for precision and coverage, and comparison of WMA index scores against two external datasets to examine if scores aligned with known indicators of housing instability.

### Results

Manual review confirmed that the PIT index effectively captured accommodation signals in case notes. WMA index scores showed strong alignment with external indicators of housing instability.

### Conclusion

This study demonstrates the feasibility of a rule-based text mining approach for identifying housing-related needs and tracking signals over time through the PIT and WMA indices. Validation results highlight the WMA index as a useful proxy for housing vulnerability, supporting its potential application in future research examining the relationship between housing and criminal justice outcomes.

---

# INTRODUCTION

---

Housing instability refers to frequent movement across different and often temporary forms of accommodation, including short-term housing and homelessness (Plage et al., 2024). It is strongly associated with a range of adverse outcomes, including poor health and disrupted access to services (Davies & Wood, 2018; Gordon et al., 2025). In criminal justice contexts, housing instability can occur rapidly in the weeks following release from custody (Taylor & Bartels, 2025). Alarming, about half of the people exiting custody expect to be in short-term or emergency accommodation, and a small percentage anticipate sleeping rough (Australian Institute of Health and Welfare, 2023). Post-release housing instability has been associated with an increased risk of reincarceration, where individuals experiencing frequent moves or homelessness are twice as likely to return to custody within nine months than those with stable housing (Baldry et al., 2003). Conversely, access to public housing has been associated with improved criminal justice outcomes including reductions in police incidents, court appearances and time spent in custody (Martin et al., 2021).

Being able to measure and monitor housing needs is critical for providing timely interventions in cases of instability and for assessing the impact of reintegration programs on post-release outcomes. Despite this importance, the ability to effectively track housing<sup>1</sup> needs among justice involved individuals can be challenging. In the context of Corrective Services NSW (CSNSW),

structured data on activities addressing individuals' housing can be drawn from two primary sources: internal CSNSW service records and linked external datasets from formal service providers such as Specialist Homelessness Services (SHS). Each source offers valuable insights but also has distinct limitations. For example, while CSNSW provides a range of post-release support and reintegration services, housing status is typically recorded based on self-reported information at points of service referral and exit. Consequently, events prior to referral, after service exit and between these discrete time points may not be captured.

On the other hand, external service datasets such as those provided by SHS can offer a more nuanced view of housing needs through multiple touch points such as at service provision and follow-up interactions. However, these datasets depend critically on individuals actively seeking and engaging with these services. This creates a significant gap in available data for those who do not engage with formal service providers, either due to stigma or mistrust, and those experiencing hidden homelessness, such as couch surfing or staying temporarily with friends or family. Furthermore, interaction with formal service providers often occur downstream in the trajectory of housing instability such as when an individual is already experiencing homelessness, which limits visibility into housing transitions and emerging risks. External datasets may also be inconsistently recorded and siloed across different service providers, making it difficult to form a unified picture of an individual's housing stability.

---

<sup>1</sup> In this paper we use the term "housing" to encompass both short-term accommodation arrangements and longer-term housing situations.

Given these limitations, the current study explores the potential of leveraging unstructured case notes as an alternative data source to more comprehensively capture housing stability among individuals who are supervised in the community by CSNSW Community Corrections. Community supervision orders involve mandated contact between an individual and their supervising officer. Each contact may include administrative updates or substantive discussions around behaviour change and other needs such as housing or employment. These interactions are documented in open-text case notes which serve as narrative records of individual circumstances intended to support case management and continuity of care.

While case notes generally serve an administrative purpose, they also represent a potential source of untapped data which may provide new opportunities to better understand and measure housing needs, including how these needs relate to post-release outcomes and other aspects of reintegration. Case notes are collected consistently across all individuals under active supervision. This allows the formation of a more complete picture of housing experiences from a broad sample of individuals, including those who are not engaged with formal service providers or who do not currently present with acute housing needs. Additionally, case notes are captured at regular intervals, providing a consistent temporal frame for observing trajectories in housing stability over time. This is particularly important because housing instability is highly dynamic, often involving rapid changes in circumstances or patterns of deterioration that can lead to homelessness (Baldry et al., 2003; Johnson et al., 2014). Regular case note data may therefore enable the

detection of emerging risks and patterns of instability that would be missed in datasets which rely on event triggered records.

To our knowledge, no prior work has systematically extracted housing-related signals from supervision case notes. This gap likely reflects the limitations of using traditional data extraction methods to access information that is recorded in unstructured formats. The recognition of the potential that case notes hold and the limitations of existing data extraction methods has driven our current approach to develop a bespoke text mining tool to identify and extract housing-related signals embedded within case notes.

This report documents the development process and validation of our text mining tool, addressing both methodological design and empirical evaluation. In selecting a technique for text analysis, we considered two broad options: rule-based systems and large language models (LLM). Rule-based systems rely on manually crafted patterns, keywords, and linguistic rules to detect relevant information, while LLMs are trained from vast amounts of text data and can be used to generalise across natural language processing (NLP) tasks. We opted to develop a bespoke rule-based tool over using LLMs. This methodological choice was shaped by several practical considerations.

While LLMs, particularly those based on transformer architectures such as BERT, represent the current state-of-the-art in NLP (Devlin et al., 2019; Wolf et al., 2020) and underpin widely used systems like ChatGPT, their application in low resource settings present notable challenges (Ramadass et al., 2024).

---

Training LLMs from scratch is a computationally intensive process, requiring access to high-end Graphics Processing Units (GPUs), large memory capacity, and extended training time (Treviso et al., 2023). While recent advances have made it possible to leverage pre-trained models that can be adapted to a wide range of tasks with less effort and computational costs, effective fine-tuning of these models depend on the availability of annotated training data (Parthasarathy et al., 2024; Szep et al., 2025), which is often limited and, in our case, non-existent.

Furthermore, although LLMs excel across a wide range of NLP tasks, their advantages are often best realised in complex, multipurpose applications such as conversational agents, abstractive summarisation or generative text production (Minaee et al., 2024). In contrast, rule-based systems require minimal computational resources and are well suited to simpler, task specific use cases like ours where the goal is to identify patterns of housing-related keywords in unstructured text. Within the criminal justice context, rule-based text mining systems have been successfully applied to similar problems. One example is the Case Quantify and Search Tool (C-QST) developed by CSNSW, which extracts and converts free-text information in case notes into structured data for program evaluation (Chong et al., 2017). Another example is the application of rule-based methods to extract information about coercive control behaviours and the types of injuries sustained by domestic violence victim survivors from NSW Police records (Kim & Karystianis, 2023; Karystianis et al., 2018).

Privacy concerns were also a key consideration. Case notes contain highly sensitive information. LLMs have been shown to retain and potentially expose sensitive information from training data under certain conditions (Lehman et al., 2021; Xiong et al., 2025; Ramakrishnan & Balaji, 2025). Additionally, publicly available LLM services such as ChatGPT may leverage user interactions to improve model performance unless disabled through privacy settings (OpenAI Help Center, 2025). While enterprise services and pre-trained LLMs typically exclude user data from use in model improvement, the default behaviour of many consumer-facing models underscores the importance of full control over data usage and retention. Although these risks are less pronounced in our context where we do not intend to deploy our system publicly, the transparency and governance afforded by an in-house rule-based system is preferred in an environment where privacy and interpretability are paramount.

---

## METHODS

### Development

The development of our text mining tool broadly involved two key steps. First, we constructed a keyword dictionary to capture references to an individual's housing situation. This dictionary incorporated rule-based patterns to account for contextual framing, enabling us to distinguish between different types of housing-related mentions and to handle linguistic cues such as negation<sup>2</sup>, and temporal context (e.g., contemporary versus historical housing status).

---

<sup>2</sup> Negation refers to identifying when a statement reverses meaning (e.g., "no housing issues" vs. "housing issues").

Second, we derived two indices to aggregate keyword matches: one to estimate housing need at the time of each record, and another to track a person's housing stability over time. All development and analysis were conducted using the R programming environment (R Core Team, 2024) and the tidyverse suite of packages for data manipulation and visualization (Wickham et al., 2019).

## Data

Case notes from individuals who transitioned from custody to parole supervision in the community between January 2015 and December 2017 were extracted to support tool development. For each individual in this cohort, only the case notes recorded during the first three months of their supervision period were included. This time-bounded selection was designed to capture an early phase of transition which is often marked by heightened housing support needs and adjustment challenges (Baldry et al., 2003; Pleggenkuhle, Huebner, & Kras, 2015). Focusing on this initial phase facilitated the extraction of meaningful keywords while reducing computational load.

It should be noted that case notes in this dataset were not limited to documentation of face-to-face supervision sessions with the supervisee. They also included administrative and other internal notes, which may capture discussions or actions related to housing needs. Including these sources ensured that the analysis reflects a

broader set of signals about accommodation-related issues, beyond direct client interactions.

The dataset was pre-processed to exclude case notes containing fewer than 20 words to reduce noise and computational load. This threshold was informed by exploratory analysis which showed that shorter notes were predominantly administrative notes and rarely contained informative content related to housing needs. This reduced the dataset from 237,424 to 202,670 case notes from 16,590 individuals. The retained notes had an average length of 148 words and a maximum of 1309 words<sup>3</sup>.

## Procedure

### Initial keyword seeding

Due to the absence of an annotated dataset, the process began with a simple heuristic approach. Case notes were scanned for the presence of four seed keywords: 'homeless', 'couch surfing', 'temporary accommodation', and 'sleeping rough'. Case notes containing any of these terms were classified as indicating housing-related needs, while those without were considered not to reflect such needs. This initial binary classification provided a foundational split for further analysis.

### Dictionary expansion

To expand the keyword set beyond the initial seed terms, a Term Frequency–Inverse Document Frequency (TF-IDF) analysis was conducted. TF-IDF is a widely used technique that is used to

---

<sup>3</sup> Due to technical limitations, each case note was limited to a maximum of 3,000 characters. About 2% of case notes exceeded this threshold and were truncated. This applies to both the development and validation datasets.

---

identify words that are particularly distinctive or informative within a subset of documents (Ramos, 2003; Manning, Raghavan, & Schütze, 2008).

In this context, TF-IDF scores were calculated by comparing term distributions between two groups of case notes: those flagged as housing-related (based on the initial seed keywords) and those not flagged. The term frequency (TF) component measures how often a word or phrase appears in a group of documents, while the inverse document frequency (IDF) assigns lower weights to terms that are common across all documents, highlighting words that are more specific to a particular group. This allowed us to identify additional terms that were statistically more prominent in housing-related notes. For example, terms like 'boarding house', 'transitional housing' or 'sleeping in car' were not part of the initial seed list but emerged as relevant through the TF-IDF analysis. This analysis also supported Named Entity Recognition (NER) by surfacing frequently mentioned names of service providers and shelter facilities. Additionally, it helped identify acronyms commonly used in case notes (e.g., 'doh' for Department of Housing). The TF-IDF analysis was conducted iteratively, with newly identified keywords added to the initial keyword seed list after each iteration. This cycle was repeated until the authors determined that no new relevant keywords were emerging.

Prior to the TF-IDF computation, case notes were converted to lower case, stripped of punctuations and tokenised into multi-word expressions using

whitespace-based splitting. Tokenisation included bigrams (two-word sequences) and trigrams (three-word sequences) to capture common phrases such as 'couch surfing' and 'in temporary accommodation'. Tokenisation into longer sequences was excluded as the added value was limited and did not justify the increase in computational load. Additionally, a customised list of stop words (common words used in the English language, e.g., 'the', 'and') were removed to reduce noise and focus the analysis on semantically meaningful terms.

To further operationalise the expanded keyword list, a set of extraction rules was developed using regular expressions (regex). Regex is preferred as it allows for targeted extraction of relevant phrases and patterns, including non-standard expressions commonly found in case notes whereas generalisation using other common methods such as stemming or lemmatisation<sup>4</sup> can introduce ambiguity and errors<sup>5</sup>. Regex also enables the matching of keyword variants, including common misspellings and abbreviations. Exclusion rules were also included to remove irrelevant matches (e.g., distinguishing between historical vs. current experience of homelessness).

The resulting keyword dictionary and processing rules were manually reviewed by the authors. Disagreements were resolved through discussion until consensus was reached. The final list of agreed keywords was then grouped into the thematic categories shown in Table 1.

---

<sup>4</sup> Stemming and lemmatisation are text normalisation techniques used in NLP to reduce individual words to their base or dictionary form. For example, words like accommodation, accommodated, accommodating can be reduced to their root form of accommodate.

<sup>5</sup> Stemming may result in matches for words like *accommodating* which may be of less relevance to the current use case.

**Table 1.** Housing-related themes captured by the text mining tool.

Theme	Description	Example keywords	N. Keywords
Homelessness	Direct references to being homeless or living rough.	Currently homeless, sleeping rough	38
Motel/Hotel	Mentions of temporary stays in commercial accommodation, often used as short-term accommodation.	Hostel, motor inn	9
Crisis Accommodation	Mentions of urgent housing provided through specialist services or shelters.	Emergency housing, crisis accommodation, overnight accommodation	21
Temporary Accommodation	Mentions of non-permanent housing arrangements, including boarding houses	Residing temporarily, boarding house,	42
Caravan	Mentions of residence in caravan parks	Caravan	2
Eviction	Mentions of forced displacement or loss of housing	Evicted	2
Housing Services	Mentions of interactions with housing service providers, including applications, or placements	SHS, Housing NSW, Link2Home	15
Rental Application	Mentions of support required to facilitate rental applications	Applying for rental, find private rental	18

## Indices construction

To summarise housing-related keyword activity, two indices were developed.

### Point-in-Time (PIT) Index

The PIT index aggregates keyword matches within each case note by summing across thematic categories. To prevent inflated scores from repeated mentions of the same concept, counts of multiple keyword hits under a single theme observed within a single case note were collapsed. This aggregation approach ensured that the index reflects the breadth of housing-related themes captured rather than the frequency of mentions.

A higher PIT score, however, does not necessarily indicate greater severity of need. For example, it may signal that more themes were discussed,

which could reflect proactive engagement rather than unmet need. Similarly, a low or absent PIT score does not necessarily indicate the absence of need. It could be that relevant concerns were not documented, or that other more pressing issues aside from housing took precedence in the indexed session. This limitation highlights the need for a complementary measure to better capture patterns and sustained needs over time.

### Weighted Moving Average (WMA) Index

The WMA index was developed to complement the limited usages of the PIT index. The WMA index is intended to produce a signal that is capable of contextualising extreme PIT scores within a broader temporal frame. Rather than treating aberrant PIT scores as definitive indicators of need, the WMA interprets them in relation to surrounding case notes, helping to

---

distinguish meaningful patterns from short-term volatility. This is conceptually important because housing instability can be fluid and unpredictable; emerging gradually at times or triggered by sudden events. Examining patterns over time helps identify whether a spike reflects a one-off disruption or part of a sustained period of vulnerability, which is critical for understanding risk and planning timely interventions.

The WMA aggregates PIT scores across a rolling window of case notes. The rolling window is set to one month, meaning each WMA score reflects PIT activity up to the preceding 30 days from the indexed case note. Custom weights were applied based on the number of days elapsed, with more recent notes contributing more heavily to the final score. This weighting reflects the assumption that recent case notes are more indicative of current or emerging needs and helps the index remain responsive to changes in an individual's circumstances. The size of the rolling window and weighting scheme are configurable and can be adjusted to suit different research or operational needs.

Together, the PIT and WMA indices provide complementary views of housing-related needs. The PIT provides a snapshot of need referenced within a single case note while the WMA index is particularly valuable for identifying emerging, sustained or escalating needs over time.

## Validation

The validation strategy addressed both content and criterion validity. Content validity was assessed through manual review of a sample of case notes to confirm that thematic keywords

were adequately captured by the PIT index. Criterion validity was examined by comparing WMA scores against two external service engagement datasets, examining whether elevated WMA scores corresponded with known periods of housing instability.

## Data

A new set of case notes was extracted to serve as an unseen dataset for validation purposes. This set of case notes were from a more recent cohort of individuals who exited custody into a community supervision episode between 2022 and 2024. This was intended to provide insights into the possible shifts in language use and case noting practices over time. As with the training data, only the first three months of case notes were extracted for each individual and the filtering parameters used to pre-process the training data were applied. This yielded a final validation dataset of 190,093 case notes from about 15,590 individuals. These case notes had an average length of about 224 words and a maximum of 1239 words.

## Content validity

The authors reviewed the extracted keywords and their source texts for a sample of 500 case notes, assessing whether thematic hits and the PIT index aligned with their manual coding. Discrepancies were discussed and thematic groupings and keyword dictionaries were refined where necessary. This exercise confirmed that the identified keywords remained relevant, and no substantial revisions were required. While a small number of keywords did not return matches within the sample, they were retained to ensure sensitivity to historical case notes and to account

for the possibility that relevant mentions may have been missed due to the limited sample size.

Although good practice involves computing inter-rater reliability statistics, this was not formally conducted because of the near perfect agreement between the authors in both the development and validation phases.

## Criterion validity

Two validation datasets were used: the CSNSW Funded Partnership Initiative (FPI) service provision dataset and Bureau of Crime Statistics and Research (BOCSAR)'s Linked Data Asset (LinDA). Both datasets document individuals' housing needs, but they differ in origin, structure, and scope, offering complementary perspectives for validation. Broadly, we expect that higher WMA index scores will correspond to greater housing-related need as recorded in these datasets. For example, individuals with more severe housing needs such as those experiencing homelessness would exhibit higher WMA scores than those in more stable housing situations, e.g., those who secured public housing. It is important to note that these datasets only capture individuals who were engaged with services. While this is a limitation which restricts the ability to assess and validate index scores for those outside of these service streams, these datasets nevertheless represent the closest available approximation to ground truth and provide a valuable benchmark for assessing the tool's performance.

### FPI

The FPI is a CSNSW initiative that provides funding to not-for-profit and government organisations delivering services that support the

management of people being supervised in the community. FPI comprises three service streams, one of which supports individuals transitioning from custody to the community in meeting their housing needs. Of relevance to this study is the administrative data collected within this stream which includes Transitional Supported Accommodation (TSA), Initial Transitional Support (ITS) and Extended Reintegration Service (ERS). These programs are all focused on housing-related support during reintegration and provide a strong basis for validating an index designed to detect housing needs. This dataset holds structured information on a person's engagement with these services and their housing status at service referral and case closure.

### LinDA

LinDA integrates information from multiple service systems, including social housing, SHS, child protection, Out of Home Care (OOHC), and criminal justice agencies. Although LinDA does not capture individuals who are entirely absent from service systems, the larger number of contributing service streams mean that LinDA may capture individuals who do not appear in FPI but are nonetheless seeking or receiving support through other formal service providers. Importantly, LinDA can also capture individuals who may not be actively engaged in support programs but are still visible in administrative systems, such as those on public housing waiting lists or on public housing tenancy. The possibility for partial visibility into unmet needs and extrapolation of housing stability between service contacts across multiple agencies makes LINDA a valuable complement to FPI, particularly for individuals who may not receive dedicated

---

housing support through CSNSW but still interact with other service systems.

### Procedure

Data tables from LinDA were received separately from each participating service provider. As a result, the first step involved grouping and arranging each individual's contacts across multiple providers into a unified chronological sequence. This required matching individuals across datasets using a common identifier and then ordering all recorded interactions by date, regardless of the originating service. This step constructs a coherent timeline of service engagement to ensure that subsequent WMA tagging and housing status alignment were based on as complete a view of each individual's history as possible.

For each documented service event, the WMA from the case note written closest in time to the event was extracted. To be eligible for selection, the case note must be written within a two-week window of the service event and can be from either before or after the event. The two-week window was selected as it aligns with operational practice, where individuals assessed as medium or higher risk of reoffending generally have supervision contact at least once a fortnight. While restricting the matching window resulted in the exclusion of some service events that are temporally distant from any case note, especially for low-risk individuals who may have less frequent contact, this trade-off was considered necessary as a narrower window better preserved the temporal relevance of the WMA index.

Where multiple case notes fell within the two-week window and were equally close to the

service timestamp, the case note preceding the event was selected. This reflects an alignment with the calculation of the WMA which is derived from case notes leading up to the indexed note. The extracted WMA score was then assessed against the individual's housing status as recorded at the time of the service event.

---

## RESULTS

### Descriptive analysis and illustrative example of tool outputs

Before presenting validation results, we first provide an overview of how the tool operates and summarise its outputs across the full validation case note dataset. Approximately 32% of case notes contained at least one accommodation-related keyword and had a PIT score greater than 0, indicating detected needs in relation to housing.

Basic descriptives are presented in Table 2 which shows the average and range of WMA and PIT scores across the entire dataset and at the end of the first, second, and third months of supervision. Across all case notes, both the WMA and PIT scores ranged from 0 to 6 (mean WMA = 0.23, mean PIT = 0.22). Both indices generally declined over time, possibly reflecting reduced housing-related needs following the immediate post-release period and as supervision progressed.

To illustrate how these indices function at the individual level, Figure 1 presents the trajectory of PIT and WMA scores for one individual, plotted against the chronological order of case notes (there were 13 case notes in total) rather than calendar dates. The two Y-axes show the WMA

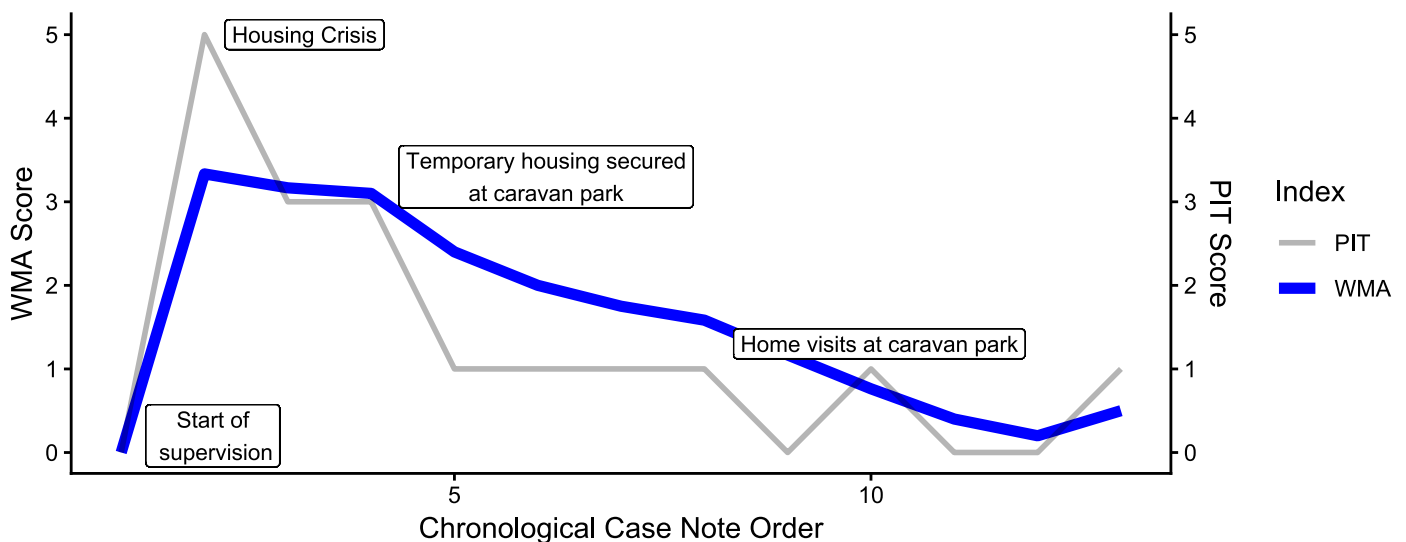
(left axis, blue line) and PIT scores (right axis, grey line) and additional annotation is included to show significant housing events that were documented in the case notes. Immediately after the start of supervision, there was a sharp spike in PIT, reflecting an acute housing crisis. Both indices declined over time as the individual secured temporary accommodation at a caravan park, but the patterns differed. PIT scores exhibited greater volatility, fluctuating sharply and occasionally reaching zero when housing was not discussed in a case note.

The WMA declined more gradually and did not reach zero, signalling a sustained yet precarious state while the individual resided in a caravan park. This smoothing effect of the WMA highlights its ability to capture ongoing need even when short-term signals disappear.

Given the PIT's volatility, it currently serves primarily as an intermediate step toward constructing the WMA index, which is our main measure of housing-related need. Accordingly, the validation presented in the following section focuses on the WMA index.

**Table 2.** Descriptive statistics of PIT and WMA scores across the entire validation dataset.

	Mean PIT (SD)	Mean WMA (SD)	Range PIT	Range WMA
Overall	0.22 (0.57)	0.23 (0.47)	0.0 – 6.0	0.0 – 6.0
End of first month	0.19 (0.52)	0.23 (0.43)	0.0 – 6.0	0.0 – 3.5
End of second month	0.15 (0.45)	0.17 (0.33)	0.0 – 4.0	0.0 – 3.0
End of third month	0.15 (0.44)	0.15 (0.32)	0.0 – 4.0	0.0 – 3.0



**Figure 1.** Single case study demonstrating tool functionality.

---

## FPI

At the time of referral to one of FPI's service providers, an individual's self-reported accommodation status is captured using one of the six categories shown in Table 3. To validate the WMA index, we extracted the WMA score from the case note closest to the referral date, effectively tagging the referral with a corresponding housing status and WMA index score.

Across the timeframe of the validation dataset, we identified case notes from 1,041 individuals who also appeared in the FPI dataset. After applying our matching parameters which retained WMA scores originating from case notes written within a two-week window of a referral, the available data was reduced to 499 individuals. Table 3 presents the average WMA index score for each accommodation category.

**Table 3.** Mean WMA index score across housing categories at FPI referral.

Housing type	N. Cases	Mean WMA
Homeless	90	0.80
Crisis/temporary accommodation	158	0.86
Social Housing	23	0.27
Living with partner/family/friend	220	0.22
Private rental	7	0.03
Own Home	1	0.00

It is important to note that WMA scores are not inherently meaningful on their own; rather, their utility lies in relative comparison across housing types. In this regard, Table 3 shows a clear gradient in mean WMA scores across housing categories. Individuals in crisis or temporary

accommodation and those who were experiencing homelessness had the highest mean WMA scores (0.86 and 0.80 respectively), indicating a stronger presence of housing-related themes in their case notes leading up to the time of referral.

In contrast, individuals living in social housing, with family, partners, or friends, or in private rental had substantially lower WMA scores (ranging from 0.03 to 0.27), reflecting more stable housing situations. However, caution should be exercised when interpreting these scores as several categories had very small sample sizes, particularly in the private rental and own home categories. For these reasons, formal statistical testing was not conducted, and the validity of the index as examined through the FPI data is supported primarily by these descriptive patterns.

## LinDA

Records of 4,410 individuals from the validation dataset were identified in LinDA. After applying our matching parameters, this was reduced to 1,923 individuals across 4,312 points of contact. Because the data originated from multiple sources, a wide range of housing descriptors (42 categories) were observed. To enable comparability, similar categories were consolidated (e.g. 'living in a tent' and 'no dwelling, in the open' were collapsed into a single category) and categories with fewer than 20 cases were omitted resulting in the 12 categories presented in Table 4.

Table 4 shows that Individuals in the most precarious circumstances such as those with no dwelling (mean WMA = 1.19), staying in hotels/motels (1.14), or emergency

accommodation (1.11) exhibited the highest WMA scores. Transitional situations, including those moving from custodial arrangements (1.07) or in temporary accommodation (0.98), also showed elevated scores. In contrast, individuals in more stable housing contexts, such as those tenanted in public housing (0.53), and rent subsidy arrangements (0.60) recorded substantially lower WMA scores.

**Table 4.** Mean WMA index score across LinDA housing categories.

Accommodation type	N. Cases	Mean WMA
No dwelling, in the open	118	1.19
Hotel/motel	54	1.14
Emergency accommodation	429	1.11
Transition from custodial arrangement	337	1.07
Temporary accommodation	2197	0.98
Housing waitlist	500	0.85
Boarding house	64	0.72
House/townhouse/flat	161	0.67
Applicant recently housed	128	0.66
Rent subsidy	46	0.60
Public housing tenancy	88	0.53
Couch surfer	86	0.53

It is important to note that individuals included in the LinDA validation dataset would all be considered to have some degree of housing need by default, given their visibility in service systems related to housing and homelessness. This aligns with the high average WMA scores observed and supports the notion that even those with relatively lower scores are likely to be in precarious housing situations rather than stable accommodation.

While a monotonic relationship between WMA index scores and housing stability is conceptually expected, it remains unclear how the severity of instability of the housing categories in Table 3 may be ranked. In the absence of such a ranking, we conducted a Kruskal-Wallis test to detect overall WMA index score differences across housing categories. As expected, the results indicated a statistically significant difference among groups,  $H(11) = 171.15$ ,  $p < .001$ , suggesting that housing type is associated with variation in WMA index scores.

Post-hoc Dunn's tests with Bonferroni correction were then conducted under the assumption that there are three broad strata of needs: (1) severe instability (i.e., no dwelling), (2) moderate instability (temporary accommodation), and (3) relative stability (public housing tenancy). The tests revealed that individuals with no dwelling had significantly higher WMA scores than all other accommodation types ( $p < .001$  for all) except for those who are in similarly precarious circumstances such as those in a hotel/motel, emergency accommodation or transitioning from custodial arrangements. In contrast, those who were in temporary accommodation had significantly higher WMA than those in more stable housing situations such as those in public housing ( $p < .001$ ).

Contrary to expectations, although couch surfing aligns conceptually with moderate or high instability, those in this living arrangement exhibited the lowest WMA, which was significantly lower than those who were in temporary accommodations ( $p < .001$ ).

---

## DISCUSSION

---

This study demonstrates the feasibility and validity of a rule-based text mining approach for identifying housing-related needs in community supervision case notes. By constructing two indices, the PIT and WMA, we provide a framework for tracking housing instability signals over time. While the tool itself does not classify or differentiate housing types, the validation results showed that higher WMA scores aligned with external indicators of housing arrangements characterised by increasing precarity and vulnerability. These findings not only support the potential of the WMA index as a proxy indicator of housing-related needs but also demonstrate the effectiveness of our text mining approach in extracting indicators from unstructured case notes.

A key strength of our text mining approach lies in its ability to provide broad coverage across the supervised population. By leveraging case notes which are routinely collected as part of supervision, the tool can serve as an index of housing-related needs for individuals who may not otherwise have structured data recorded in this area; for example, those who are not engaged with formal service providers or have active case management goals related to accommodation. From a research perspective, this capability offers a valuable opportunity to support multiple lines of inquiry, including understanding need and patterns of instability post-release. For example, it allows the generation of comparison cohorts for evaluations which could help researchers conduct more inclusive assessments of housing-related

interventions and examine how patterns of housing instability evolve over time and are associated with criminal justice outcomes. Beyond research, the WMA index has potential operational applications where it could be deployed as a near real-time monitoring tool. Its ability to flag emerging risks and trends could support frontline staff working with supervisees, enabling timely interventions.

Flexibility is another important strength of the system. Many aspects of the tool were developed in a modular format and can be tailored to specific operational or research needs. For example, the library of keywords can be expanded, and the rolling windows and weighting schemes used to generate the WMA index can be configured. Modular design is widely recognised as good practice in NLP system development because it allows systems to be maintainable and adaptable. One example of this principle in practice is the General Architecture for Text Engineering (GATE), a widely used suite of tools for building NLP applications, which introduced a structured framework for combining modular text-processing components into configurable pipelines (Bontcheva et al., 2002). By adopting similar principles, our approach ensures our tool remains transparent and responsive to different research and operational needs.

While the WMA index captures the frequency and concentration of housing-related mentions and is intended to reflect degree of severity of housing needs, the index could be made more informative and sensitive by applying weights to themes. Although certain themes, such as homelessness, may indicate greater or more acute instability than rental support, the index currently treats all thematic hits equally. Future work could explore

applying differential weights, so that a mention of homelessness contributes more to the index than a reference to rental application support. Similarly, the tool could be extended to classify accommodation types, enabling direct identification of housing contexts from case notes. Although thematic hits can offer some classification of housing contexts, refinement would be needed to capture signals of stability. At present, stability is inferred indirectly through the absence of instability related mentions, rather than through explicit detection of stability indicators. Expanding the keyword set to include indicators such as secure housing, living with family, or home ownership could support this goal. These additions were not implemented in the current tool due to the variability and context-dependence of stability related language which is more complex to detect.

Some limitations are noted. First, validation was performed using datasets from external service providers and as a result, the WMA index was primarily evaluated on individuals who are known to formal services and present with some level of instability. While our capacity to evaluate the tool was constrained by the scope of available data, this did not diminish the ability of the tool itself to identify instability signals across the broader population. Second, the validation was conducted on a three-month observation window which was chosen to balance data availability and computational feasibility. This is, however, a relatively short period of time and it remains an open question whether the tool can fully capture longer-term housing trajectories. We have no strong reason to doubt its ability to extend to longer periods, but future iterations could test whether expanding the observation window may

improve detection of sustained patterns or transitions in housing stability.

Third, some housing arrangements reflect a form of hidden homelessness, where individuals lack secure or stable housing but do not necessarily present this as a concern during supervision. Couch surfing or temporarily staying with friends or acquaintances without formal tenancy is a common example (Homelessness Australia, 2023). Research indicates that couch surfing is associated with significant housing instability and elevated risks, including poor mental health and limited social support, although is often underreported (Curry et al., 2017; Hail-Jares et al., 2020; Mosher et al., 2025). Because of this underrepresentation, individuals who couch surf may receive deceptively low WMA scores. This is not a limitation of our tool per se but rather reflects an inherent constraint of the case noting process, where informal living arrangements are less likely to be raised or documented. While this underscores the need for caution when interpreting results for vulnerable populations who may be experiencing hidden homelessness, it also highlights an opportunity to improve operational practices by encouraging more systematic documentation of informal housing arrangements in case notes, which could enhance both service delivery and future analytical efforts.

These limitations also point to important directions for future work. Improving documentation practices is one step, but the potential of this approach extends beyond operational refinements. One of the key considerations that led us to adopt a rule-based approach was the complete lack of annotated data, which makes training more advanced

---

models impossible at this stage. The outputs from our tool such as keyword matches and index scores can serve as interpretable labels for creating annotated datasets, laying the foundation for more sophisticated methods such as the training of in-house machine learning models or LLMs. These methods can perform similar tasks at scale and with greater advantages in their ability to capture more nuanced signals that rule-based systems may miss.

More broadly, this work represents the first known attempt to systematically extract housing-related signals from community supervision case notes. It also serves as a proof of concept for leveraging case notes to measure and monitor other areas of need or social capital, such as dynamics of employment or financial stability, experienced by people in the community. These extensions could enable not only a richer understanding of the challenges faced by individuals under supervision, but also a more nuanced view of how different contextual factors interact with criminal justice outcomes. Such insights could inform both targeted interventions and broader policy strategies aimed at reducing recidivism.

## REFERENCES

---

- Australian Institute of Health and Welfare. (2023). The health of people in Australia's prisons 2022 (Cat. no. PHE 334). Australian Government.  
<https://www.aihw.gov.au/reports/prisoners/the-health-of-people-in-australias-prisons-2022/contents/summary>
- Baldry, E., McDonnell, D., Maplestone, P., & Peeters, M. (2003). Ex-prisoners and accommodation: what bearing do different forms of housing have on social reintegration?  
[https://www.ahuri.edu.au/sites/default/files/migration/documents/AHURI\\_Final\\_Report\\_No46\\_Ex\\_prisoners\\_and\\_accommodation.pdf](https://www.ahuri.edu.au/sites/default/files/migration/documents/AHURI_Final_Report_No46_Ex_prisoners_and_accommodation.pdf)
- Bontcheva, K., Cunningham, H., Maynard, D., Tablan, V., & Saggion, H. (2002). Developing reusable and robust language processing components for information systems using GATE. In *Proceedings. 13th International Workshop on Database and Expert Systems Applications* (pp. 223-227). IEEE.  
<https://ieeexplore.ieee.org/abstract/document/1045902>
- Chong, C. S., Raudino, A., Thaler, O., & Howard, M. (2017). The Case Quantify and Search Tool (C-QST). Research Publication 56, Corrective Services NSW.  
<https://correctiveservices.dcj.nsw.gov.au/documents/research-and-statistics/research-publications/rp56-case-quality-and-search-tool.pdf>
- Curry, S. R., Morton, M., Matjasko, J. L., Dworsky, A., Samuels, G. M., & Schlueter, D. (2017). Youth homelessness and vulnerability: How does couch surfing fit? *American Journal of Community Psychology*, 60(1-2), 17-24.  
<https://onlinelibrary.wiley.com/doi/abs/10.1002/ajcp.12156>
- Davies, A., & Wood, L. J. (2018). Homeless health care: meeting the challenges of providing primary care. *Medical Journal of Australia*, 209(5), 230-234.  
[https://www.mja.com.au/system/files/issues/209\\_05/10.5694mja17.01264.pdf](https://www.mja.com.au/system/files/issues/209_05/10.5694mja17.01264.pdf)
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT* (pp. 4171-4186).  
<https://doi.org/10.48550/arXiv.1810.04805>
- Gordon, S. J., Baker, N., Marin, T. S., & Steffens, M. (2025). Health Deficits Among People Experiencing Homelessness in an Australian Capital City: An Observational Study. *International Journal of Environmental Research and Public Health*, 22(2), 135.  
<https://www.mdpi.com/1660-4601/22/2/135>
- Hail-Jares, K., Vichta-Ohlsen, R., & Nash, C. (2021). Safer inside? Comparing the experiences and risks faced by young people who couch-surf and sleep rough. *Journal of Youth Studies*, 24(3), 305-322.  
[https://brisyouth.org/wp-content/uploads/2021/06/Safer-Inside\\_HailJares\\_VichtaOhlsen\\_Nash\\_2020-2.pdf](https://brisyouth.org/wp-content/uploads/2021/06/Safer-Inside_HailJares_VichtaOhlsen_Nash_2020-2.pdf)
- Homelessness Australia. (2023). Homelessness in Australia: Facts and figures. Retrieved from  
<https://homelessnessaustralia.org.au>
- Johnson, G., Moschion, J., Scutella, R., Tseng, Y. P., & Wooden, M. (2014). Journeys Home research report no. 3. Findings from Waves 1 to 4 – Development of a Homelessness Typology. Melbourne Institute of Applied Economic and Social Research.

- [https://melbourneinstitute.unimelb.edu.au/\\_data/assets/pdf\\_file/0007/2202856/Johnson\\_et\\_al\\_Journeys\\_Home\\_Research\\_Report\\_W3\\_Reissue.pdf](https://melbourneinstitute.unimelb.edu.au/_data/assets/pdf_file/0007/2202856/Johnson_et_al_Journeys_Home_Research_Report_W3_Reissue.pdf)
- Karystianis, G., Adily, A., Schofield, P., Knight, L., Galdon, C., Greenberg, D., ... & Butler, T. (2018). Automatic extraction of mental health disorders from domestic violence police narratives: text mining study. *Journal of medical internet research*, 20(9), e11548.  
<https://www.jmir.org/2018/9/e11548/>
- Kim, M. T., & Karystianis, G. (2023). Text mining police narratives of domestic violence events to identify coercive control behaviours. *Crime and Justice Bulletin*, 260.  
<https://bocsar.nsw.gov.au/documents/publications/cjb/cjb251-300/cjb260-report-coercive-control-in-domestic-violence.pdf>
- Lehman, E., Jain, S., Pichotta, K., Goldberg, Y., & Wallace, B. C. (2021). Does BERT pretrained on clinical notes reveal sensitive data?. <https://doi.org/10.48550/arXiv.2104.07762>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press. <https://doi.org/10.1017/CBO9780511809071>
- Martin, C., Reeve, R., McCausland, R., Baldry, E., Burton, P., White, R. and Thomas, S. (2021) Exiting prison with complex support needs: the role of housing assistance, AHURI Final Report No. 361, Australian Housing and Urban Research Institute Limited, Melbourne,  
<https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=3913797>.
- Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large language models: A survey. arXiv preprint [arXiv:2402.06196](https://arxiv.org/abs/2402.06196).
- Mosher, H. I., Weeks, M. R., & Li, J. (2023). "Call Back When You Are More Homeless:" Missed Targets for Prevention and Risk Reduction Among Youth Who Are Couch Surfing. *International Journal on Homelessness*, 1-30.  
<https://ojs.lib.uwo.ca/index.php/ijoh/article/download/18666/17780>
- OpenAI Help Center. (2025). How OpenAI uses your data. <https://help.openai.com/en/articles/7730893-how-openai-uses-your-data>
- Parthasarathy, V. B., Zafar, A., Khan, A., & Shahid, A. (2024). The ultimate guide to fine-tuning llms from basics to breakthroughs: An exhaustive review of technologies, research, best practices, applied research challenges and opportunities. arXiv preprint [arXiv:2408.13296](https://arxiv.org/abs/2408.13296).
- Plage, S., Stambe, R.-M., Parsell, C., & Kuskoff, E. (2024). Climbing, stalling, falling: How people experiencing housing instability anticipate their futures. *Journal of Sociology*, 61(1), 40-61.  
<https://doi.org/10.1177/14407833241255151>
- Pleggenkuhle, B., Huebner, B. M., & Kras, K. R. (2015). Solid start: Supportive housing, social support, and reentry transitions. *Journal of Crime and Justice*, 39(3), 380-397.  
<https://www.tandfonline.com/doi/abs/10.1080/0735648X.2015.1047465>
- R Core Team. (2024). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.
- Ramadass, K., Singh, A. P., & Kalyani, S. (2024). Textless NLP-Zero Resource Challenge with Low Resource Compute. arXiv preprint [arXiv:2409.19015](https://arxiv.org/abs/2409.19015).
- Ramakrishnan, A., & Balaji, S. (2025). Assessing and mitigating data memorization risks in fine-tuned LLMs. <https://doi.org/10.48550/arXiv.2508.14062>
- Ramos, J. (2003). Using TF-IDF to determine word relevance in document queries. *Proceedings of the First Instructional Conference on Machine Learning*.  
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b3bf6373ff41a115197cb5b30e57830c16130c2c>
- Szep, M., Rueckert, D., von Eisenhart-Rothe, R., & Hinterwimmer, F. (2024). A practical guide to fine-tuning language models with limited data. arXiv preprint [arXiv:2411.09539](https://arxiv.org/abs/2411.09539).
- Taylor, H., & Bartels, L. (2025). 'Somewhere We Can Call Home and... Be Normal': Findings From the Justice Housing Programme Evaluation. *Australian Journal of Social Issues*.  
<https://openresearch-repository.anu.edu.au/bitstreams/1a858a2b-1b70-479e-a660-6a870e080bb2/download>
- Treviso, M., Lee, J. U., Ji, T., Aken, B. V., Cao, Q., Ciosici, M. R., Hassid, M., Heafield, K., Hooker, S., Raffel, C., Martins, P. H., Martins, A. F., Forde, J. Z., Milder, P., Simpson, E., Slonim, N., Dodge, J., Strubell, E., Balasubramaniam, N., ... Schwartz, R. (2023). Efficient methods for natural language processing: A survey. *Transactions of the Association for Computational Linguistics*, 11, 826-860.  
<https://aclanthology.org/2023.tacl-1.48.pdf>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of open source software*, 4(43), 1686.  
<https://joss.theoj.org/papers/10.21105/joss.01686>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., ... Rush, A. M. (2020).

---

Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations* (pp. 38-45).

<https://aclanthology.org/2020.emnlp-demos.6.pdf>

Xiong, Y., Zhang, T., & Chen, M. (2025). The landscape of memorization in large language models.

<https://doi.org/10.48550/arXiv.2507.05578>

## Other CRES Research Titles

- Dec 2025 Reducing reoffending through intervention diversity: Evidence from the Intervention Pathways model
- Dec 2025 The relationship between participation in intervention pathways and short-term reoffending
- Dec 2025 Exploring the relationship between intervention dosage and reoffending: Evidence from an integrated intervention delivery model in NSW correctional centres
- Nov 2025 Exploring perceptions of correctional climate among people on community supervision in New South Wales
- Oct 2025 Environmental influences on perceptions of correctional climate: A multilevel modelling analysis
- Sept 25 Evaluation of the Practice Guide for Intervention (PGI): Assessments of quality assurance and supervisee outcomes
- Aug 2025 Identification of criminogenic needs using the PARRCC assessment tool
- July 2025 Profiles of criminogenic need among Aboriginal and non-Aboriginal men and women in prison
- June 2025 Conducting survey research using inmate digital tablets: Are respondents representative of the inmate population?
- Dec 2024 Evaluation of the Practice Guide for Intervention (PGI): Triangulating perceptions of dual role relationships and their associations with staff practice and supervisee outcomes
- Nov 2024 Exploring perceptions of correctional climate among people in prison in New South Wales
- Nov 2024 Exploring perceptions of organisational fairness and safety attitudes among correctional staff in NSW
- Sept 2024 Implementing digital technologies in Prison: Inmates' ongoing experiences of tablet access and connections with family and friends
- Sept 2024 Drivers of program participation in NSW correctional centres: A trend analysis of digital tablets and Five Minute Interventions (FMI)
- June 2024 Full body x-ray scanners at NSW correctional centres: Scanner activity, contraband, and behavioural outcomes
- June 2024 Full body x-ray scanners at NSW correctional centres: Staff and inmate perspectives
- April 2024 Evaluation of throughput and participation outcomes for the Intervention Pathways model
- March 2024 Evaluation of the Compulsory Drug Treatment Program: Within-treatment change
- Feb 2024 Examination of treatment dosage delivered through the Intervention Pathways Model
- Dec 2023 Assessment and experience of prison climate in NSW correctional centres
- Dec 2023 Five Minute Intervention (FMI) skills acquisition by correctional staff: The role of manager buy-in
- Dec 2023 Dynamics of wellbeing and needs satisfaction among people in prison
- Dec 2023 Empirical review of the Pathways to Employment (P2E) pilot for women in prison



Research Bulletin No.71  
ISSN 2207 0850  
© Corrective Services NSW

Corrections Research, Evaluation & Statistics  
Delivery, Performance & Culture  
Locked Bag 5000  
Parramatta NSW Australia 2124  
Email:  
[research.enquiries@correctiveservices.nsw.gov.au](mailto:research.enquiries@correctiveservices.nsw.gov.au)