# Mathematical Methodology for Defining a Frequent Attender within Emergency Departments

**Elizabeth Williams** [1,2,*], **Syaribah N. Brice** [1,2] **and Dave Price** [2]

[1] *School of Mathematics, Cardiff University, Cardiff, United Kingdom*
[2] *Digital & Health Intelligence, Cardiff and Vale University Health Board, Cardiff, United Kingdom*

Correspondence*:
Corresponding Author
WilliamsEM20@Cardiff.ac.uk

## ABSTRACT

**Objective:** Emergency department (ED) frequent attenders (FA) have been the subject of discussion in many countries. This group of patients have contributed to the high expenses of health services and strained capacity in the department. Studies related to ED FAs aim to describe the characteristics of patients such as demographic and socioeconomic factors. The analysis may explore the relationship between these factors and multiple patient visits. However, the definition used for classifying patients varies across studies. While most studies used frequency of attendance to define the FA, the derivation of the frequency is not clear.
**Methods:** We propose a mathematical methodology to define the time interval between ED returns for classifying FAs. K-means clustering and the Elbow method were used to identify suitable FA definitions. Recursive clustering on the smallest time interval cluster created a new, smaller cluster and formal FA definition.
**Results:** Applied to a case study dataset of approximately 336,000 ED attendances, this framework can consistently and effectively identify FAs across EDs. Based on our data, a FA is defined as a patient with three or more attendances within sequential 21-day periods.
**Conclusion:** This study introduces a standardised framework for defining ED FAs, providing a consistent and effective means of identification across different EDs. Furthermore, the methodology can be used to identify patients who are at risk of becoming a FA. This allows for the implementation of targeted interventions aimed at reducing the number of future attendances.

**Keywords: Emergency Department, Frequent Attender, K-means Clustering, Health Services, Health Care Utilisation, Targeted Interventions**

## 1 INTRODUCTION

Emergency departments (EDs) are the first point of contact for patients who require urgent medical care. The ED is a complex system influenced by many factors, including the patient, the staff, the hospital, and the community. The ED is a dynamic system that is constantly changing. The number of patients attending EDs has increased over the years, with attendances in England rising by 20% between 2008 and 2018 (Budhwani et al., 2022). This increase in attendances has resulted in EDs becoming overcrowded, with patients experiencing long waiting times. The increase in attendance has also led to a rise in the number frequent attenders (FAs). FAs, sometimes referred to as frequent flyers or frequent users, are patients who attend EDs multiple times within a given period. Although this group of patients represents a small minority, they contribute significantly to the volume of ED attendances, which continues to increase each year (Hunt et al., 2006; Pek et al., 2022; Jacob et al., 2016).

It is important to recognise that the high utilisation of ED services by FAs is not simply a reflection of these patients being a burden on the healthcare system. Rather, FAs are often dealing with complex, chronic medical and social issues that are not being adequately addressed elsewhere in the care system. Conditions such as asthma (Wakefield et al., 1997), cancer (Wong et al., 2018), neurological issues (Lago et al., 2019) and mental health and alcoholism concerns (Jacob et al., 2016; Mak et al., 2022) contribute significantly to individuals making multiple visits to the ED.

ED FAs are associated with an increased risk of admission, mortality (Greenfield et al., 2020), and hospital length of stay (LOS) (Street et al., 2018). Ultimately, they contribute to the high utilisation of ED personnel or other services such as ambulance services (Wooden et al., 2009; Shen et al., 2018), primary care (Williams et al., 2001), community health centres (Savageau et al., 2006), and social services (Byrne et al., 2003). The challenge is to identify these patients systematically and provide them with coordinated, comprehensive care.

Studies related to FAs aim to investigate the factors that influence repeated visits, the broader impact of FA attendances on health and social services and the efficacy of intervention programmes in mitigating attendances. These influencing factors encompass a spectrum, ranging from the clinical challenges faced by patients to their demographic profiles, environmental context and behavioural patterns. While exploring factors associated with FAs is a prevalent focus in the literature, it is worthwhile noting that the definition of an FA can vary. The most commonly employed criterion for identifying FAs involves their attendance frequency. Typically, a FA is defined as a patient who attends the ED more than a determined threshold within a given period. Table 1 demonstrates the different definitions used in the literature across different countries. Within these papers, there was no justification for how they defined a FA. When different definitions are employed within the same country, a significant challenge arises in determining the true prevalence of FAs. Recognising and accurately identifying FAs is critical because they are known to contribute to the high expenses of health services (Furia et al., 2023; Ruger et al., 2004; Williams, 2001). It is important to recognise that these individuals have needs that require attention, even if those needs are not clearly linked to a specific medical condition (Neal et al., 2000). This means they may face ongoing challenges in their daily lives, such as social, emotional or psychological issues, that do not always show up in traditional medical assessments but significantly impact their overall well-being (Birrenbach et al., 2022).

The definition of FAs should be tailored to each ED, driven by data and reflective of the unique patient characteristics and demographics of each ED. Acknowledging the variability across different EDs, a data-driven approach ensures a more relevant understanding of FAs within each setting. Furthermore, the definition should be dynamic enough to operate in real-time, allowing the prompt identification of individuals who are at risk of becoming FAs. This real-time capability is crucial for swiftly implementing targeted interventions aimed at reducing the number of future attendances by ensuring the needs of these individuals are met.

By aligning the definition with the specific attributes of each ED's patient population and incorporating a real-time capability, healthcare providers can enhance their ability to proactively address the complex medical, social and environmental factors that put certain individuals at risk of becoming FAs (Al-Jaroodi et al., 2020). Rather than simply viewing FAs as a challenge posed by the patients themselves, this approach recognises it as a reflection of shortcomings in the broader system of care (Ablard et al., 2017). With this understanding, providers can work to implement holistic, patient-centred interventions that connect FAs with the appropriate resources and support they need.

This study aims to introduce a mathematical methodology for determining the time interval between successive visits to EDs that qualifies a patient as a FA. This proposed methodology holds the potential to establish a standardised framework applicable not only to the specific ED under investigation but also transferable for defining FAs in other EDs. The distinctive advantage of employing a mathematical approach lies in its precision and adaptability. Unlike alternative methods, a mathematical methodology offers a systematic and data-driven way to objectively identify FAs based on time intervals between visits. Furthermore, this method's versatility enhances its applicability to a wide range of EDs, accommodating the variability in patient demographics and characteristics. By offering a standardised yet flexible approach, the proposed mathematical methodology provides a valuable tool for healthcare providers seeking a consistent and effective means of identifying FAs.

## 2 METHODS

### 2.1 Study Design and Setting

This study employed a retrospective cohort design, utilising data from a major ED department in Wales, UK. Spanning a six-year period, from January 2017 to December 2022, the research sought to explore patterns of ED visits and define FAs based on time intervals between return visits. The retrospective cohort design allowed for the examination of historical data to discern trends and associations.

### 2.2 Participant Selection

The study population consisted of adult patients, defined as individuals aged 17 years or older, who sought medical attention at the ED between January 2017 and December 2022, who were classified as a 'Major' patient. The ED defines a major patient as one who requires very urgent emergency care (NHS England, 2024). The decision to focus exclusively on 'Major' patients in this study was guided by the need to align the model with the clinical and operational priorities of EDs. Major patients are characterised by high-acuity conditions that necessitate significant resources and urgent care. Their frequent attendances impose a disproportionate burden on ED resources, making them a critical group for targeted intervention strategies (Iacobucci, 2022). However, it is important to note that this classification is inherently subjective and may vary across institutions. By tailoring the model to identify patterns specific to this cohort, the study aims to address the most pressing challenges associated with ED overcrowding and resource allocation.

Our initial dataset contained 336,898 patients. We applied data preprocessing to retain only patients with complete demographic and clinical information relevant to our analysis. First, 66 patients were removed due to missing National Health Service (NHS) numbers. Next, 233 patients lacked recorded admission or discharge dates and times, leading to their exclusion. We also filtered out 512 patients based on implausible age data, specifically those recorded as being over 120 years old. Finally, 46 patients were removed due to incomplete information on gender.

This filtering, designed with the specific purpose of enabling the ability to track patients and time between visits, resulted in a total of 184,051 patients, contributing to a total of 336,041 ED visits. Of these visits, 152,183 (45%) were identified as return visits, denoting subsequent visits made after the initial identified visit. Notably, 34.6% of patients made more than one visit within the specified six-year timeframe. This comprehensive dataset provided a robust foundation for the subsequent analyses.

## 2.3 Statistical Analysis

This research introduces an innovative approach to defining FAs by shifting from traditional frequency-based definitions to a time interval-based definition. Our research investigated diverse statistical clustering methodologies to establish an enhanced definition of frequent ED users that goes beyond simple visit counts. While previous research typically relied on basic frequency measurements to classify FAs, we sought to develop a more sophisticated approach by considering both the number of visits and the temporal spacing between them. Although Monte Carlo methods, which are well-known in healthcare management for their probabilistic sampling approach, might seem an appropriate choice given their familiarity in ED settings, we opted for a deterministic clustering approach. Unlike Monte Carlo simulations, which provide probabilistic approximations that vary between runs, our selected methodology needed to deliver consistent, reproducible classifications of frequent attenders. This requirement for deterministic outcomes, combined with our focus on identifying distinct patterns in visit frequency and timing, led us to explore various clustering techniques.

We systematically evaluated multiple clustering algorithms to determine their effectiveness. Our approach aligns with machine learning model selection principles, including data preparation, feature selection and model optimisation, as detailed by Ramlakhan et al. (2022b), to ensure robust and reproducible outcomes. Initially, we investigated hierarchical clustering through the agglomerative approach, which demonstrated strengths in gradual cluster formation but proved impractical for our extensive ED dataset due to its computational demands. We then explored Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and the Ordering Points to Identify the Clustering Structure (OPTICS) algorithms, which excel at identifying irregularly shaped clusters and isolating outliers. However, these methods presented challenges in our context due to their requirement for specific density parameters, which proved difficult to optimise given the variable patterns of ED utilisation.

We also considered Gaussian Mixture Models (GMM) for their probabilistic approach to cluster assignment. Yet, the underlying assumption of normally distributed data made this method suboptimal for our ED visit patterns, which typically follow non-Gaussian distributions. The K-modes algorithm was also evaluated, particularly for its strength in categorical data analysis, but it failed to adequately capture the temporal aspects of visit patterns that were crucial to our study.

After comprehensive testing, K-means clustering initially emerged as the optimal methodology. However, due to the wide range of time scales present in the data, we extended this to a two-stage K-means approach. The initial K-means method was selected based on several factors: its computational efficiency when handling large datasets, its ability to effectively process both visit frequency and inter-visit intervals and its straightforward implementation approach. To determine the ideal number of clusters for each stage, we employed the Elbow method, which helped us identify the point where additional clusters provided minimal additional benefit, thus optimising the balance between model complexity and accuracy.

The final framework, combining two-stage K-means clustering with Elbow method optimization, represents a significant advancement in identifying ED frequent attenders. This hierarchical approach provides a more nuanced understanding compared to traditional counting methods by first grouping patients based on their temporal visit patterns, followed by sub-clustering based on visit frequencies. Our methodology offers a more comprehensive and accurate way to identify and understand patterns in ED utilisation, particularly when dealing with varying time scales, which can better inform healthcare resource allocation and intervention strategies.

### 2.3.1 K-means Clustering

In the broader literature, K-means clustering has been successfully used on ED data to predict patient outcomes and ED utilisation. For example, Grant et al. (2020) applied K-means to identify patients with complex profiles, predicting diverse healthcare utilisation and mortality outcomes. Additionally, Liu et al. (2017) used K-means clustering to create a triage system within EDs, with Huang et al. (2008) clustering patients against medical utilisation, discovering that their FA population, more often utilised other medical services.

To provide insight into the mathematical model for the clustering algorithm, the K-means method clusters together ED patient data points with similar returns to ED times. K-means clustering is an iterative machine learning algorithm designed to partition a dataset into $K$ clusters, where each data point is assigned to the cluster with the nearest mean. The goal is to minimise the within-cluster sum of squares, which is represented by the objective function:

$$J = \sum_{i=1}^{K} \sum_{j=1}^{n} \left\| x_j^i - c_i \right\|^2 \tag{1}$$

where $n$ is the number of data points, $x_j^i$ is the $j^{th}$ data point in cluster $i$ and $c_i$ is the centroid of cluster $i$. $\left\| x_j^i - c_i \right\|^2$ is the Euclidean distance between the data point and the centroid.

To provide a definition for FA, recursive clustering was applied to the cluster with the smallest LOS mean ($n_1$). This means the K-means clustering algorithm was then performed again on the smallest cluster to create $m$ new clusters from cluster $n_1$. The cluster with the smallest LOS mean ($m_1$) was then used to define a FA. This method is known as two-stage K-means clustering (Salman et al., 2011). In two-stage clustering, the initial clustering identifies broad groups in the data. Then, clustering is applied a second time to divide the groups into more granular segments. Applying this technique to healthcare data has provided useful insights in many cases (Ayanore et al., 2016; You-Shyang et al., 2012; Marshman et al., 2016). For more detail about the K-means method, we have provided a step-by-step applied example in the supplementary material.

### 2.3.2 Elbow Method

Determining the optimal number of clusters is an important step in the K-means clustering method. The Elbow method is a heuristic method used to determine the optimal number of clusters ($K$) in a dataset. The idea is to perform the K-means algorithm for a range of values $K$, calculate the sum of squared distances from each point to its assigned centre ($J$), and plot the results. The 'elbow' point is often considered the point where the rate of decrease of $J$ slows down, suggesting that the addition of more clusters does not significantly reduce the within-cluster sum of squares. The algorithm uses the same objective function as the K-means clustering algorithm, Equation (1). The Elbow method offers an empirical approach to determine the appropriate number of clusters based on the data itself. However, it is important to note that identifying the exact location of the elbow involves a qualitative assessment and subjective judgement.

#### 2.3.3 Scenario Impact

This study examines the impact of FAs on ED resources by quantifying 'hours of patient ED time', representing the cumulative hours FAs spend in EDs. This metric serves as an indicator of the system burden created by FAs. We calculated this based on the average duration each patient remained in the ED per visit. Estimating 'hours of patient ED time' allows for a clearer assessment of how reducing specific frequent attendance patterns could alleviate ED capacity strains, aligning with the NHS four-hour wait target for patients.

## 3 RESULTS

This section will present the results of applying the K-means clustering and Elbow method to the ED data. The results discussed will include the descriptive statistics of the data, the clustering results, and the definition of a FA.

### 3.1 Descriptive Statistics

The K-means clustering and Elbow method as discussed previously was performed on the six years' worth of data from 1st January 2017 to 31st December 2022. The total number of patients included in the dataset was 184,051 who made a total of 336,041 visits. The proportion of male and female patients is 48% and 52% respectively. The distribution of male and female patients' ages is significantly different (Kruskal Wallis test = 201.9, p-value $< 0.001$). The mean age for male patients is 49 years old (Q1 = 30, Q3 = 66) compared to the mean age for female patients of 48.6 years old (Q1 = 27, Q3 = 69). The average total time spent in the ED department was calculated to be 10.7 hours (sd = 16.8 hours) per patient. However, the median total time spent was found to be 5.6 hours with 50% of patients staying in the ED between three and 11 hours. At an aggregated level, there was no significant difference between male and female patients' time spent in the ED (p-value = 0.25). Within 30 days of attending the ED, 25% of patients will present again. Around 36.2% female patients have return visits compared to 33.1% male patients.

### 3.2 Clustering Results

The Elbow method was performed on the data, with the results showing the elbow occurs at $K = 3$. Therefore, the K-means clustering algorithm was performed with $K = 3$. The results show three determined clusters between 0 and 324 days, 325 and 913 days and, 914 and 2171 days (Table 2). The smallest cluster is between 0 and 324 days, with 68.58% of patients falling into this cluster. The mean time between visits for this cluster is approximately three months.

Since returning within 324 days provides too broad of a definition for a FA, the K-means clustering algorithm was performed again on this cluster. Taking the smallest cluster of returning between 0 and 324 days, the Elbow method was performed again to determine the optimal number of clusters. The results show an Elbow method of eight clusters. Therefore, performing the K-means algorithm with eight clusters results in the smallest cluster being between 0 and 21 days. Overall, there are 31.4% of patients falling into this cluster. The mean time between visits for this cluster is 7.08 days (Table 3). This cluster accounts for 9.72% of all attendances. In practice, this represents 10.1% of patients. In our analysis, we found that age did not significantly impact the clustering results, indicating that the observed patterns in patient characteristics were consistent across both adult and elderly populations.

### 3.3 Definition of a Frequent Attender

The results from the two-stage K-means clustering algorithm show that the smallest cluster is between 0 and 21 days. In order to provide a formal definition, this cluster can be examined further.

Our provisional definition will specify a threshold number of visits within each 21 day period as qualifying for FA status. As illustrated in Figure 1 the 21 day count is reset after each visit, so the frequency pattern is evaluated over rolling 21 day windows. This approach quantifies FA based on the number of admissions within sequential 21 day periods, rather than just looking at the initial admissions. By scrutinising the admission patterns in this manner, we can precisely define FAs as exhibiting a high number of admissions in 21 days, recurrently over time. This definition captures the essence of multiple frequent admissions over successive short intervals, which is the key characteristic of FAs.

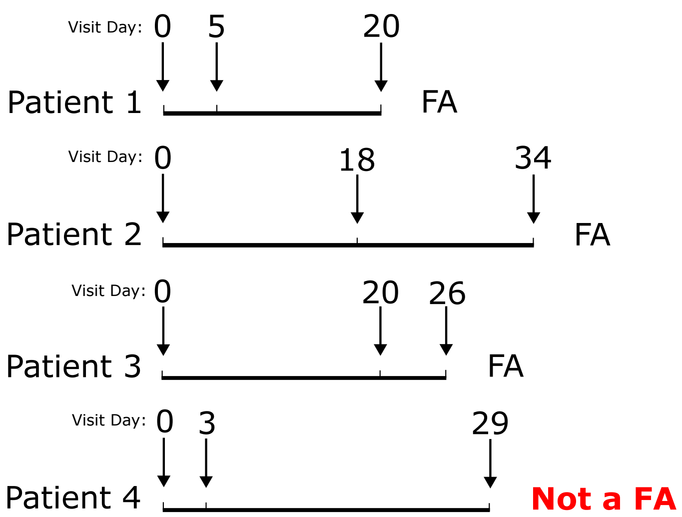


**Figure 1.** Visualisation of FA definition

Analysis of the 0 to 21 day cluster, reveals patients in this group exhibit a high rate of recurrent readmissions over short time intervals. Specifically, on average, if a patient in this cluster attends within 21 days of discharge, they will have an additional 1.8 attendances in the subsequent 21 days. As depicted within Figure 1, there is a pattern of multiple repeat admissions within 21 day subsequential periods. Therefore, based on the ED data used, we propose a FA is defined as a patient who has three or more attendances within sequential 21 day periods.

The selection of the 21-day period was guided by the results of K-means clustering, which identified this as the smallest cluster with frequent and recurrent attendances. This interval aligns with acute care utilisation patterns observed in FAs and represents a meaningful timeframe for intervention. The choice of three attendances within the 21-day window reflects a balance between sensitivity and specificity. While two attendances in this period may signify episodic or transient patterns, three or more visits indicate sustained and potentially preventable reliance on ED services. This threshold ensures the definition targets a group whose frequent usage imposes significant demands on resources.

It is acknowledged that patients attending less frequently, such as once per month, may not meet the FA definition if they only accumulate two attendances in any successive 21-day clusters. However, this reflects an intentional focus on patients with more intense usage patterns, who are most likely to benefit from targeted care interventions. Those with less frequent but consistent visits are typically better served by outpatient or primary care pathways rather than interventions aimed at reducing high-frequency ED usage.

# 4 DISCUSSION

This study introduced a novel application of K-means clustering and the Elbow method to define frequent ED attenders based on time intervals, instead of visit frequency. Applied to six years of data from one major ED, the technique found a pattern of repeat visits within 21 day periods. This identified a data driven definition of a FA as a patient who has three or more attendances within sequential 21 day periods, capturing recurrent high utilisation of ED services over short intervals. When compared to the literature, as shown in Table 1, the 9% FA rate is comparable with other studies, demonstrating the validity of the approach. The proposed methodology offers a standardised framework to quantify FAs in an adaptive way which can be tailored to each ED's unique patient population.

Using the suggested framework, the definition for a FA is three attendances within sequential 21-day periods. This definition is data-driven and tailored to the specific ED dataset analysed. However, applying the framework to different ED data would likely result in a different definition, reflecting the variability in patient demographics and attendance patterns across settings. While it is challenging to directly compare frequency-based methods with time-interval approaches, our methodology offers greater precision by identifying FAs more quickly. By focusing on shorter intervals, it can detect changes in attendance patterns, including seasonal variations, more quickly. This allows for earlier identification of FAs and enables the timely implementation of targeted interventions, which is crucial for managing the high resource demands associated with frequent attendance.

## 4.1 Generalisability

While this study introduced a methodology using data from a single ED, the overall mathematical approach has the potential for broader generalisability. The technique provides a standardised framework that could be applied to other EDs to adaptively define FAs based on their specific patients and utilisation patterns. However, the generalisability may be influenced by factors such as patient demographics, healthcare system differences and access to primary care.

The optimal clustering patterns and FA definitions may differ across EDs, with substantially different contexts and patient mixes. Therefore, while the method itself is generalisable, with a structured framework provided, the specific resulting FA definition would likely need to be tailored and validated for each ED environment based on its unique characteristics. This includes testing the model on unseen datasets and validating against domain-expert-defined FA populations to ensure its applicability across settings (Graham et al., 2018; Ramlakhan et al., 2022a; Yao et al., 2021). This flexible mathematical approach means an adaptive solution can be implemented, rather than imposing a universal standard definition that fails to capture local variability. Additional research across diverse settings is needed to further test the methodology and refine definitions suited to different ED contexts.

## 4.2 Scenario Impact

The impact of FAs on EDs is substantial, with high utilisation leading to increased waiting times, costs and reduced quality of care (Hoot and Aronsky, 2008; Moskop et al., 2009; Pines et al., 2011). However, it is crucial that this is done in a compassionate, patient-centred manner that recognises the complex needs of this population, rather than simply viewing them as a burden on the system. While the 21-day period is used to identify recurrent patterns of ED use and forecast immediate resource savings, it is important to emphasise that interventions targeting FAs are not expected to yield results within such a short timeframe. The true value of identifying FAs early lies in the ability to implement long-term, coordinated care strategies, such as chronic disease management and mental health support, that can help reduce persistent ED use over time. The 21-day window serves as a tool for early identification, enabling proactive and sustained care that addresses the root causes of frequent ED visits.

If the fourth visit within 21 days is prevented, this would result in 43,075.20 hours of patient ED time saved per year, with an average of 6.3 hours per patient in an ED. If this was able to be increased to prevent third visits onwards within 21 days, this would result in 62,937 hours of patient ED time saved per year.

Calculating the average time patients spend with an ED doctor within our dataset equates to 42 minutes per patient. This is comparable to other ED studies (Walker et al., 2019; Wrede et al., 2020). This value can be used to estimate the impact of FAs on ED resources. For example, if we prevent visits from the fourth attendance within 22 days, this would result in 3,418.67 minutes of doctor time saved per year. This equates to 142.44 hours per year. Similarly, if we were to reduce this to the third visit, this would increase to 208.13 hours of doctor time saved per year. By saving doctor hours, this could be used to treat other patients, reducing waiting times, improving the quality of care and reduce the pressures faced on current staff.

It is important to note that the goal of this work is not to simply reduce or restrict FA attendance, as that could further exacerbate the challenges they face in accessing appropriate care. Rather, the aim is to use the identification of FAs as an opportunity to proactively content them with the coordinated, patient-centred services they require to address the root cause driving their high utilisation. This compassionate, system-level approach has the potential to improve outcomes for this vulnerable population by supporting them to access care and support from providers suitable for their needs, whilst also alleviating pressures on the emergency care system.

### 4.3 Limitations

This study has some limitations to consider. Firstly, the data was used from a single ED from a single country, which might not be fully representative of other EDs or healthcare systems. Secondly, even though six years' worth of data was used, this may not have been sufficient to capture the full range of utilisation patterns. Our data included the period of Covid19 when traditionally there were fewer attendances to ED (Lateef, 2020). This could have resulted in patients not being classified as a FA when they would have been in a non-Covid period. Thirdly, the use of the 'Major' patient classification in the ED is inherently subjective and can vary across different institutions. The categorisation of patients as 'Major' or 'Minor' typically relies on clinical judgment, which introduces variability that could impact the generalisability of the model. Another limitation is that the data was limited to the ED and did not include other healthcare services. This means that patients who were classified as a FA may have been using other healthcare services, which could have resulted in a different FA definition. Finally, the study only focused on temporal patterns in defining FAs, whereas with additional clinical and demographic factors, further insight could have been provided into high utilisers of ED services.

## 5 CONCLUSION

In conclusion, this study presents an innovative methodology for defining FAs based on time intervals between ED visits. The approach offers a standardised framework that can be applied to different EDs to define FAs based on their specific patients and utilisation patterns. This methodology provides a data driven approach to precisely identify FAs, enabling targeted interventions to reduce future attendances. By shifting from traditional frequency based definitions to a time interval based definition, a FA can be identified more efficiently. This mathematical approach provides a valuable tool for healthcare providers seeking a consistent and effective means of identifying FAs. Future research should focus on utilising the method with clinical and demographic factors to further refine FA definitions and explore the impact in practice of interventions on FAs.

## 6 ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements

## 7 CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## 8 FUNDING

The research reported in this publication was supported by the collaboration between Cardiff University and Cardiff and Vale University Health Board, with funding gratefully received from the Digital and Health Intelligence directorate.

## 9 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the invaluable insights from the staff of the Emergency Department at Cardiff and Vale University Health Board. Specifically, we would like to express our sincere thanks to Anna Sussex, former National Programme Manager for Frequent Attenders within Public Health Wales and founder of the Welsh Emergency Department Frequent Attender Network, for sharing her knowledge and insights into the ED Frequent Attender programme.

## REFERENCES

Ablard, S., Coates, E., Cooper, C., Parry, G., and Mason, S. M. (2017). Can more appropriate support and services be provided for people who attend the emergency department frequently? National Health Service staff views. *Emergency Medicine Journal* 34

Al-Jaroodi, J., Mohamed, N., and Abukhousa, E. (2020). Health 4.0: On the Way to Realizing the Healthcare of the Future. *IEEE Accesss* 8, 211189 – 211210

Ayanore, M., Pavlova, M., and Groot, W. (2016). Focused maternity care in Ghana: results of a cluster analysis. *BMC Health Services Research* 16

Birrenbach, T., Hoffmann, M., Hautz, S. C., Kämmer, J. E., Exadaktylos, A. K., Sauter, T. C., et al. (2022). Frequency and predictors of unspecific medical diagnoses in the emergency department: a prospective observational study. *BMC Emergency Medicine* 22, 1–10

Budhwani, N., Capanna, M., Baban, A., Isetta, M., Oldman, A., Daniels, N., et al. (2022). Care planning or case management for frequent emergency department attendance in adults. *Cochrane Database of Systematic reviews* 15

Butler, A., Love, A., Young, J., and Kinner, S. (2020). Frequent attendance to the emergency department after release from prison: a prospective data linkage study. *Journal of Behavioral Health Services and Research* 47, 544–559

Byrne, M., Murphy, A., Plunkett, P., McGee, H., Murray, A., and Bury, G. (2003). Frequent attenders to an emergency department: A study of primary health care use, medical profile, and psychosocial characteristics. *Annals of Emergency Medicine* 41, 309–318

Chan, J.-E., Tin, A., Chow, W., Tiah, L., Tiru, M., and Lee, C. (2018). Frequent attenders at the emergency department: an analysis of characteristics and utilisation trends. *Proceedings of Singapore Healthcare* 27, 12–19

Cordell, B., Nguyen, K., Shahir, A., Lord, S., and Gallego, G. (2022). Prevalence and characteristics of frequent presenters to Auburn Hospital emergency department. *EMA - Emergency Medicine Australasia* 34, 898–906

Furia, G., Vinci, A., Colamesta, V., Papini, P., Grossi, A., Cammalleri, V., et al. (2023). Appropriateness of frequent use of emergency departments: A retrospective analysis in Rome, Italy. *Frontiers in Public Health*

Graham, B., Bond, R., Quinn, M., and Mulvenna, M. (2018). Using Data Mining to Predict Hospital Admissions From the Emergency Department. *IEEE Access* 6, 10458–10469

Grant, R. W., McCloskey, J., Hatfield, M., Uratsu, C., Ralston, J. D., Bayliss, E., et al. (2020). Use of Latent Class Analysis and k-Means Clustering to Identify Complex Patient Profiles. *JAMA Network Open* 3, e2029068–e2029068

Greenfield, G., Blair, M., Aylin, P., Saxena, S., Majeed, A., Hoffman, M., et al. (2020). Frequent attendances at emergency departments in England. *Emergency Medicine Journal* 37, 597–599

Greenfield, G., Blair, M., Aylin, P. P., Saxena, S., Majeed, A., and Bottle, A. (2021). Characteristics of frequent paediatric users of emergency departments in England: An observational study using routine national data. *Emergency Medicine Journal* 38, 146–150

Halcomb, E., Smyth, E., and Ghosh, A. (2017). Local trends in older people presenting at an emergency department. *Collegian* 24, 381–385

Hansagi, H., Olsson, M., Sjöberg, S., Tomson, Y., and Göransson, S. (2001). Frequent use of the hospital emergency department is indicative of high use of other health care services. *Annals of Emergency Medicine* 37, 561–567

Hoot, N. R. and Aronsky, D. (2008). Systematic review of emergency department crowding: causes, effects, and solutions. *Annals of Emergency Medicine* 52, 126–136

Hotham, R., O'Keeffe, C., Stone, T., Mason, S., and Burton, C. (2022). Heterogeneity of reasons for attendance in frequent attenders of emergency departments and its relationship to future attendance. *Emergency Medicine Journal* 39, 10–15

Huang, J.-A., Weng, R.-H., Lai, C.-S., and Hu, J.-S. (2008). Exploring Medical Utilization Patterns of Emergency Department Users. *Journal of the Formosan Medical Association* 107, 119–128

Hunt, K. A., Weber, E. J., Showstack, J. A., Colby, D. C., and Callaham, M. L. (2006). Characteristics of frequent users of emergency departments. *Annals of emergency medicine* 48, 18–20

Iacobucci, G. (2022). Poor emergency department performance is down to sicker patients, report finds. *BMJ* 356, j1130

Jacob, R., Wong, M., Hayhurst, C., Watson, P., and Morrison, C. (2016). Designing services for frequent attenders to the emergency department: A characterisation of this population to inform service design. *Clinical Medicine, Journal of the Royal College of Physicians of London* 16, 325–329

Jelinek, G., Jiwa, M., Gibson, N., and Lynch, A.-M. (2008). Frequent attenders at emergency departments: A linked-data population study of adult patients. *Medical Journal of Australia* 189, 552–556

Lago, L., Westley-Wise, V., Mullan, J., Lambert, K., Zingel, R., Carrigan, T., et al. (2019). Here one year, gone the next? Investigating persistence of frequent emergency department attendance: A retrospective study in Australia. *BMJ Open* 9

Lateef, F. (2020). The impact of the covid 19 pandemic on emergency department attendance: What seems to be keeping the patients away? *Journal of Emergencies, Trauma, and Shock* 13, 246–251

Liu, W., Wang, Z., Liu, X., Yue, W., and Bell, D. (2017). A clustering approach to triage categorization in A&E departments. In *2017 23rd International Conference on Automation and Computing (ICAC)*. 1–6

Locker, T., Baston, S., Mason, S., and Nicholl, J. (2007). Defining frequent use of an urban emergency department. *Emergency Medicine Journal* 24, 398–401

Mak, C. C. M., Mao, D. R. H., Siddiqui, F. J., Lim, A., Davamoni-Thomas, J., Tang, J. P., et al. (2022). A new paradigm in management of frequent attenders to emergency departments with severe alcohol use disorder—A pilot study for assertive community treatment in Singapore. *Frontiers in Health Services*

Marshman, Z., Broomhead, T., Rodd, H. D., Jones, K., Burke, D., and Baker, S. R. (2016). Who attends a Children's Hospital Emergency Department for dental reasons? A two-step cluster analysis approach. *Community Dentistry and Oral Epidemiology* 45, 49–58

Michelen, W., Martinez, J., Lee, A., and Wheeler, D. P. (2006). Reducing frequent flyer emergency department visits. *Journal of Health Care for the Poor and Underserved* 17, 59–69

Moore, L., Deehan, A., Seed, P., and Jones, R. (2009). Characteristics of frequent attenders in an emergency department: Analysis of 1-year attendance data. *Emergency Medicine Journal* 26, 263–267

Moskop, J. C., Sklar, D. P., Geiderman, J. M., Schears, R. M., and Bookman, K. J. (2009). Emergency department crowding, part 1—concept, causes, and moral consequences. *Annals of Emergency Medicine* 53, 605–611

Neal, R. D., Heywood, P. L., and Morley, S. (2000). 'I always seem to be there' - a qualitative study of frequent attenders. *British Journal of General Practice* 50, 716–723

NHS England (2024). Guidance for emergency departments: initial assessment. Available Online at: `https://www.england.nhs.uk/guidance-for-emergency-departments-initial-assessment/` (Accessed on 7 January 2025)

Palmer, E., Leblanc-Duchin, D., Murray, J., and Atkinson, P. (2014). Emergency department use: Is frequent use associated with a lack of primary care provider? *Canadian Family Physician* 60, e223–e229

Paul, P., Heng, B., Seow, E., Molina, J., and Tay, S. (2010). Predictors of frequent attenders of emergency department at an acute general hospital in Singapore. *Emergency Medicine Journal* 27, 843–848

Pek, P., Lau, C., Sim, X., Tan, K., Mao, D., Liu, Z., et al. (2022). Nationwide study of the characteristics of frequent attenders with multiple emergency department attendance patterns. *Annals of the Academy of Medicine* 51, 483–492

Pines, J. M., Hilton, J. A., Weber, E. J., Alkemade, A. J., Al Shabanah, H., Anderson, P. D., et al. (2011). International perspectives on emergency department crowding. *Academic Emergency Medicine* 18, 1358–1370

Quilty, S., Shannon, G., Yao, A., Sargent, W., and McVeigh, M. (2016). Factors contributing to frequent attendance to the emergency department of a remote Northern Territory hospital. *Medical Journal of Australia* 204, 111.e1–111.e7

Ramlakhan, S., Saatchi, R., Sabir, L., Singh, Y., Hughes, R., Shobayo, O., et al. (2022a). Understanding and interpreting artificial intelligence, machine learning and deep learning in Emergency Medicine. *Emergency Medicine Journal* 39, 380–385

Ramlakhan, S. L., Saatchi, R., Sabir, L., Ventour, D., Shobayo, O., Hughes, R., et al. (2022b). Building artificial intelligence and machine learning models : a primer for emergency physicians. *Emergency Medicine Journal* 39, e1

Ruger, J. P., Richter, C. J., Spitznagel, E. L., and Lewis, L. M. (2004). Analysis of costs, length of stay, and utilization of emergency department services by frequent users: Implications for health policy. *Academic Emergency Medicine* , 1311–1317

Salman, R., Kecman, V., Li, Q., Strack, R., and Test, E. (2011). Two-stage clustering with k-means algorithm. In *Recent Trends in Wireless and Mobile Networks*, ed. H. Springer, Berlin

Sandoval, E., Smith, S., Walter, J., S-AH, S., Olson, M., Striefler, R., et al. (2010). A Comparison of Frequent and Infrequent Visitors to an Urban Emergency Department. *The Journal of Emergency Medicine* 38, 115–121

Savageau, J. A., McLoughlin, M., Ursan, A., Bai, Y., Collins, M., and Cashman, S. B. (2006). Characteristics of frequent attenders at a community health center. *Journal of the American Board of Family Medicine: JABFM* 19, 265–275

Scheiner, N., Cohen, S., Davis, R., Gale, T., and Agyare, A. (2019). The effect of integrated care on self-management and emergency department attendance. *BJPsych Bulletin* 43, 117–122

Shen, Y., Tay, Y., Teo, E., Liu, N., Lam, S., and Ong, M. (2021). Association between the elderly frequent attender to the emergency department and 30-day mortality: A retrospective study over 10 years. *World Journal of Emergency Medicine* 9, 20–25

Shen, Y., Teo, E., Liu, N., Lam, S., and Hock, M. (2018). Data-driven approach to defining the emergency department frequent attender using a cohort of 10 years. *Journal of Acute Medicine* 8, 6–16

Skinner, J., Carter, L., and Haxton, C. (2009). Case management of patients who frequently present to a Scottish emergency department. *Emergency Medicine Journal* 26, 103–105

Sousa, S., Hilder, T., Burdess, C., and Bolton, P. (2019). 2.5 years on: What are the effects of a 'frequent attenders' service in the Emergency Department? *BJPsych Bulletin* 43, 112–116

Street, M., Berry, D., and Considine, J. (2018). Frequent use of emergency departments by older people: A comparative cohort study of characteristics and outcomes. *International Journal for Quality in Health Care* 30, 624–629

Sun, B., Burstin, H., and Brennan, T. (2003). Predictors and Outcomes of Frequent Emergency Department Users. *Academic Emergency Medicine* 10, 320–328

Uí Bhroin, S., Kinahan, J., and Murphy, A. (2019). Profiling frequent attenders at an inner city emergency department. *Irish Journal of Medical Science* 188, 1013–1019

Wakefield, M., Staugas, R., Ruffin, R., Campbell, D., Beilby, J., and McCaul, K. (1997). Risk factors for repeat attendance at hospital emergency departments among adults and children with asthma. *Australian and New Zealand Journal of Medicine* 27, 277–284

Walker, K., Ben-Meir, M., Dunlop, W., Rosler, R., West, A., O'Connor, G., et al. (2019). Impact of scribes on emergency medicine doctors' productivity and patient throughput: multicentre randomised trial. *BMJ* l121

Williams, E. R. L., Guthrie, E., Mackway-Jones, K., James, M., Tomenson, B., Eastham, J., et al. (2001). Psychiatric status, somatisation, and health care utilization of frequent attenders at the emergency department: A comparison with routine attenders. *Journal of Psychosomatic Research* 50, 161–167

Williams, R. M. (2001). Frequent emergency department use in sweden: implications for emergency medicine in the united states. *Annals of Emergency Medicine* , 627–629

Wong, T., Lau, Z., Ong, W., Tan, K., Wong, Y., Farid, M., et al. (2018). Cancer patients as frequent attenders in emergency departments: A national cohort study. *Cancer Medicine* 7, 4434–4446

Wooden, M. D., Air, T. M., Schrader, G. D., Wieland, B., and Goldney, R. D. (2009). Frequent attenders with mental disorders at a general hospital emergency department. *EMA - Emergency Medicine Australasia* 21, 191–195

Wrede, J., Wrede, H., and Behringer, W. (2020). Emergency Department Mean Physician Time per Patient and Workload Predictors ED-MPTPP. *Journal of Clinical Medicine* 9, 3725

Yao, L. H., Leung, K. C., Tsai, C. L., Huang, C. H., and Fu, L. C. (2021). A Novel Deep Learning–Based System for Triage in the Emergency Department Using Electronic Medical Records: Retrospective Cohort Study. *Journal of Medical Internet Research* 23, e27008

You-Shyang, C., Ching-Hsue, C., Chien-Jung, L., Cheng-Yi, H., and Han-Jhou, S. (2012). Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment. *Computers in Biology and Medicine* 42, 213–221

Zhou, J., Cameron, P., Dipnall, J., Shih, K., and Cheng, I. (2022). Using network analyses to characterise Australian and Canadian frequent attenders to the emergency department. *Emergency Medicine Australasia*

**Table 1.** Published definitions of frequent attenders of EDs.

| Authors | Country | Definition | Proportion of FA attendance |
|---|---|---|---|
| Wooden et al. (2009) | Australia | $\geq$ 1 in 1 month | 4.5% |
| Wakefield et al. (1997) | Australia | $\geq$ 2 in 12 months | 40-60% |
| Butler et al. (2020); Cordell et al. (2022); Halcomb et al. (2017); Street et al. (2018) | Australia | $\geq$ 4 in 12 months | 3.7%-64.7% |
| Jelinek et al. (2008) | Australia | $\geq$ 5 in 12 months | 2.4% |
| Quilty et al. (2016) | Australia | $\geq$ 6 in 12 months | 13% |
| Lago et al. (2019) | Australia | $\geq$ 7 in 12 months | 2.6% |
| Zhou et al. (2022) | Australia & Canada | $\geq$ 4 in 12 months | 4.7%-5.6% |
| Palmer et al. (2014) | Canada | $\geq$ 4 in 12 months | 11.3% |
| Greenfield et al. (2020) | England | $\geq$ 3 in 12 months | 27.1% |
| Greenfield et al. (2021); Locker et al. (2007); Moore et al. (2009) | England | $\geq$ 4 in 12 months | 9.1%-13.9% |
| Hotham et al. (2022); Jacob et al. (2016) | England | $\geq$ 5 in 12 months | 4.4%-11.5% |
| Williams et al. (2001) | England | $\geq$ 7 in 12 months | N/A |
| Scheiner et al. (2019); Sousa et al. (2019) | England | $\geq$ 15 in 12 months | N/A |
| Uí Bhroin et al. (2019) | Ireland | $\geq$ 3 in 12 months | 29% |
| Byrne et al. (2003) | Ireland | $\geq$ 4 in 12 months | N/A |
| Skinner et al. (2009) | Scotland | $\geq$ 10 in 6 months | N/A |
| Shen et al. (2021, 2018) | Singapore | $\geq$ 3 in 12 months | 8%-22.1% |
| Pek et al. (2022); Wong et al. (2018) | Singapore | $\geq$ 4 in 12 months | 19.6%-35.4% |
| Chan et al. (2018); Paul et al. (2010) | Singapore | $\geq$ 5 in 12 months | 8%-14.6% |
| Hansagi et al. (2001) | Sweden | $\geq$ 4 in 12 months, | 4% |
| Michelen et al. (2006) | US | $\geq$ 3 in 6 months | N/A |
| Sandoval et al. (2010) | US | $\geq$ 3 in 12 months | 7% |
| Hunt et al. (2006); Sun et al. (2003) | US | $\geq$ 4 in 12 months | 8%-28% |

| Cluster | Patient Count | % of Total Patients | Mean | Minimum | Maximum |
|---|---|---|---|---|---|
| 0 | 104359 | 68.58% | 92.32 | 0 | 324 |
| 1 | 34227 | 22.49% | 166.04 | 325 | 913 |
| 2 | 13597 | 8.93% | 1271.89 | 914 | 2171 |

**Table 2.** Results from the K-means clustering algorithm.

| Cluster | Patient Count | % of Total Patients | Mean | Minimum | Maximum |
|---|---|---|---|---|---|
| 0 | 32748 | 31.4% | 7.08 | 0 | 21 |
| 1 | 16500 | 15.8% | 35.75 | 22 | 52 |
| 2 | 12709 | 12.2% | 69.62 | 53 | 88 |
| 3 | 11038 | 10.6% | 108.13 | 89 | 129 |
| 4 | 9529 | 9.1% | 150.96 | 130 | 174 |
| 5 | 8395 | 8.0% | 197.53 | 175 | 222 |
| 6 | 7157 | 6.9% | 246.82 | 223 | 272 |
| 7 | 6283 | 6.0% | 297.79 | 273 | 324 |

**Table 3.** Results from the two-stage K-means clustering algorithm.