This version of the article may not completely replicate the final authoritative version published in Crisis: The Journal of Crisis Intervention and suicide Prevention at https://doi.org/10.1027/0227-5910/a000930.

Reference:

Lee, S. C., Rouquette, O. Y., Hawton, K., Cleobury, L., Spencer, S., Lloyd, K., Gunnell, D., Scourfield, J., & John, A. (2023). Understanding Suicide Clusters Through Exploring Self-Harm Behaviors. *Crisis*. https://doi.org/10.1027/0227-5910/a000930

Title

Understanding Suicide Clusters Through Exploring Self Harm Behaviors: a 10-year data-linkage cohort follow-up study of a Suicide Cluster using the Secure Anonymised Information Linkage (SAIL) Databank

Sze Chim Lee[1], Olivier Y. Rouquette[1], Keith Hawton[2,3], Louise Cleobury[1] , Sarah Spencer[1], Keith Lloyd[1], David Gunnell[4,5] , Jonathan Scourfield[6] , Ann John[1*]

[1]Swansea University Medical School, Swansea University, UK.

[2]Centre for Suicide Research, University of Oxford, UK.

[3]Oxford Health NHS Foundation Trust, UK.

[4] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.

[5]National Institute of Health and Care Research Biomedical Research Centre, University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol, Bristol, UK.

[6]Children's Social Care Research and Development Centre (CASCADE), School of Social Sciences, Cardiff University, UK.

Corresponding author*:

Ann John: a.john@swansea.ac.uk

Adress: Swansea University Medical School: Singleton Park, Swansea SA2 8PP, UK

Abstract

Background

There is little information about characteristics and long-term outcomes of individuals who self-harm during a suicide cluster.

Aims

To compare characteristics of individuals who self-harmed during a suicide cluster in South Wales (~10 deaths between Dec 2007 and Mar 2008) with others who self-harmed prior to the cluster, and to evaluate 10-year self-harm and mortality outcomes.

Method

Using records from the hospital serving the catchment area of the suicide cluster, enhanced by national routinely collected linked data, we created two groups: individuals who self-harmed a) during the suicide cluster, and b) one year before. We compared individuals' characteristics and performed logistic regression to compute odds ratios of 10-year self-harm and mortality outcomes.

Results

Individuals who self-harmed during the cluster were less likely to be hospitalized or have a mental health history than those who self-harmed prior to the cluster. No significant group differences were found for 10-year self-harm outcomes, but all-cause mortality was higher for males.

Limitations

Sample size was small, and data were lacking on psychological and social proximity to individuals who died during the suicide cluster.

Conclusion

Our findings highlight the importance of long-term healthcare follow-up of those who self-harm during a suicide cluster, particularly males.

Keyword:

self-harm; suicide; suicide cluster; data linkage; mortality

Abbreviations

BC – Before the Cluster

CI – Confidence Interval

DC – During the Cluster

ED – Emergency Department

ESM – Electronic Supplementary Material

ICD – International Classification of Diseases

LAA – Local Authority Area

NHS – National Health Service

OR – Odds Ratio

SAD – SAD PERSONS score

SAIL – Secure Anonymised Information Linkage

VIF – Variance Inflation Factor

Introduction

Although relatively uncommon, suicides may occur in clusters, particularly in young people (Haw et al., 2013). There are two main types of clusters described in the literature, namely, mass clusters and point clusters. While for mass clusters, often associated with media reporting of the death of a celebrity, suicide rates increase across a population within a time period, point clusters involve a concentration of suicide deaths within time and a specific locality (Joiner, 1999). There is no doubt that suicide clusters generate high levels of community distress and often widespread media attention (Hawton et al., 2015).

Several non-mutually exclusive mechanisms have been proposed underlying the initiation and maintenance of suicide clusters (Haw et al., 2013; Hawton et al., 2020). The social transmission mechanism suggests that exposure to the suicide of a significant other increases vulnerability to further suicide via imitation and suggestion or projective and pathological identification (Marchant et al., 2020). Underlying the descriptive norms is the more prevalent suicidal behavior is perceived to be, the more normalised it becomes. The assortative relating theory (Joiner, 1999; Robinson et al., 2016) proposes that the clustering of suicide is explained primarily by a group of individuals sharing certain risk factors who associate with each other and the social integration and relating mechanism refers to the effect of close-knit social networks in disseminating news and beliefs about suicides in a locality.

Nonetheless, little is known about the characteristics and long-term outcomes of those who self-harm during a suicide cluster (Haw et al., 2013). A recent qualitative study of individuals presenting with near-fatal self-harm during a suicide cluster suggested that the negative impact of the cluster could have long-term effects (John et al., 2022). We aimed to compare characteristics and long-term self-harm and mortality outcomes for individuals who self-harmed during a point cluster, with an estimated 10 deaths, which occurred in South Wales, UK, between December 2007 and March 2008 in young people aged 15-34 years (Jones et al., 2013) with those who self-harmed prior to it. This cluster was highly publicised locally and nationally by media, with a high volume of sensational reporting throughout the cluster (John et al., 2016; Marchant et al., 2020).

Methods

Study design and participants

This was a retrospective data linkage cohort study (RECORD checklist in Electronic Supplementary Material (ESM) 1) based in the Local Authority Area (LAA; population

140,000) of a suicide cluster (December 27, 2007-March 17, 2008). We used paper-based emergency department (ED) records (Suppl. Methods in ESM 2) from the district general hospital serving the locality and privacy protected routinely collected data for the Wales population from the Secure Anonymised Information Linkage (SAIL) Databank (www.saildatabank.com).

We derived two groups for this study where each group included individuals who self-harmed during the period where the suicide cluster occurred (DC group) and those who self-harmed during the corresponding period one year before (BC group). We excluded individuals who self-harmed during both periods, i.e., excluding individuals in both BC and DC groups.

ED dataset

This dataset consisted of individuals who presented to the ED of the district hospital following self-harm (index self-harm) between December 27, 2006 and March 17, 2008 by hand screening for any mention of self-harm (Suppl. Methods in ESM 2). These were then converted to electronic data by researchers for quantitative analysis. We compared characteristics and outcomes of individuals ascertained during the suicide cluster, between December 27, 2007 and March 17, 2008 (DC group, Suppl. Fig. 1 in ESM 3), with those ascertained between December 27, 2006 and March 17, 2008 (BC group, Suppl. Fig. 1 in ESM 3).

Enhanced dataset

We used routinely collected data from SAIL databank covering the Wales population between January 01, 2000 and March 16, 2018 (Suppl. Fig. 1 in ESM 3). Within the two ascertainment periods (DC and BC), we identified individuals who resided in the LLA or presented to health services located in the LAA with self-harm (primary care and hospital admission data). These individuals and those from the ED dataset were combined creating enhanced DC and BC groups (Suppl. Fig. 1 in ESM 3). Long-term outcomes were assessed by following the enhanced datasets for 10 years, starting from the date of the index self-harm event (Fig. 1A).

Data Linkage

Data from the ED dataset were uploaded to the SAIL databank, a databank that contains anonymised privacy protecting person-based linkable data from healthcare and public settings (Ford et al., 2009; Lyons et al., 2009). All data linkage was handled in accordance

with the Data Protection Act 2018 and disclosure control methods were used to restrict the reporting of small numbers (categories containing <5 individuals and related categories leading to secondary disclosure) to protect vulnerable individuals. Data between database were linked by identity matching and creation of unique anonymised linking field via a trusted organisation mandated to hold personally identifiable data. Data encryption using deterministic matching was based on National Health Service (NHS) number or probabilistic matching using available demographics (Ford et al., 2009; Lyons et al., 2009). For probabilistic linkage, a matching score was calculated to reflect the odds of matches of demographic variables for an individual. We included individuals whose data were either deterministically linked or probabilistically linked with matching score of ≥0.9. Using the matching criteria, overall accuracies of ≥99.8% could be attained and ≥94.1% of the records could be successfully linked (Lyons et al., 2009).

We used the following SAIL datasets to link the ED dataset at individual level and to identify individuals for the enhanced dataset: Welsh Demographic Service, General Practice Database, Patient Episode Database for Wales and deaths register from Office for National Statistics. Descriptions of each dataset are summarised in Suppl. Table 1 in ESM 3.

Measures

Self-harm, suicide risk, and mortality outcome

Data for current and history of self-harm, suicide attempts, and 'suicide risk' measured by the modified SAD PERSONS (SAD) score (Patterson et al., 1983) were extracted from individuals' ED record. Self-harm events and methods (categorized into overdose/poisoning, hanging/strangulation, cutting, and others/unknown) were also extracted from the primary and secondary care SAIL datasets based on previously used Read and International Classification of Diseases (ICD) version 10 codes (Marchant, Turner, et al., 2020). We extracted mortality data using ICD-10 codes and classified cause of death into all-cause, natural, unnatural, and suicide as described previously (John et al., 2018).

Other covariates

For the ED dataset, we included: sex, age, marital and household status, area deprivation as proxied by the Welsh Index of Multiple Deprivation, and urban/rural indicator. For the enhanced dataset, the same variables were used, except marital and household status (unavailable in the SAIL Databank). Other variables included physical comorbidity, previous self-harm, mental health diagnoses, alcohol and drug use, and prescription of psychotropic and opiate medications (see details in Suppl. Methods in ESM 2). These variables were

included based on previous studies on suicide and premature mortality following self-harm (Carr et al., 2017; John et al., 2020).

Statistical analysis

Full descriptions of the statistical methods are summarized in Suppl. Method (ESM 2). In brief, we compared descriptive statistics of individuals' characteristics, self-harm mortality outcomes between DC and BC groups with 95% confidence intervals (CIs). Due to small sample size, Fisher's exact tests, likelihood ratio tests and Bayes factors were used to estimate independence of variables for all contingency tables. Effect modification of stratified cross-tabulation by sex and age was tested by the homogeneity of odds ratios and Firth logistic regression model, independent sample *t* test and the associated Bayes factors were used to compare group means for continuous variables.

For the enhanced dataset, we performed univariable and multivariable Firth logistic regressions to evaluate the odds ratios (ORs) on the long-term mortality outcomes. The use of Firth regression was to circumvent the small sample bias due to small size and separation issues (Firth, 1993; Heinze & Schemper, 2002). For reference, we also presented results from conventional logistic regression for all adjusted analyses. For all adjusted analyses, we performed diagnostic checks on multicollinearity using the variance inflation factors (VIFs) of all independent variables. VIF >3 was used as a threshold of presence of multicollinearity (Miles & Shevlin, 2001).

Ethical Approval

Ethical approval was obtained from Southwest Wales NHS Local Research Ethics Committee (reference 15/WA/0366) and the Swansea University Information Governance Review Panel (reference 0319).

Results

Cohort characteristics

496 individuals were identified in ED records during December 27, 2006-March 17, 2008 and data for 402 individuals (81.0% out of 496) were successfully linked to the SAIL databank (Suppl. Fig. 1 in ESM 3). Among the 129 individuals (32.1% out of 402) who self-harmed either during the suicide cluster (DC) or during the same period a year before (BC), 86 individuals (66.7% out of 129) were from the DC and 43 (33.3%) from the BC group. From SAIL, we identified 424 additional individuals to form the enhanced dataset (N = 489) with 280 (57.3% out of 489) in the DC and 209 (42.7%) in BC group. Only <5 and 17 individuals

were excluded from the ED (<2% out of 129) and enhanced datasets (3.5% out of 489) respectively as they were ascertained in both DC and BC groups (Suppl. Fig. 1 in ESM 3).

There was no statistical evidence of differences in sociodemographic, SAD scores, and clinical characteristics between the DC and BC groups of the ED dataset (Suppl. Table 1-4 in ESM 3). However, fewer individuals in the DC group were admitted to a general or psychiatric hospital following self-harm, 7.0% (out of 86; 95% CI: 2.9%-15.1%) vs. 32.6% (out of 43; 95% CI: 19.5%-48.7%).

Sociodemographic and clinical characteristics in the enhanced DC and BC groups were similar (Suppl. Table 5-8 in ESM 3). Fewer individuals in the enhanced DC group were hospitalized with self-harm, 20.0% (out of 280; 95% CI: 15.6%-25.3%) vs. 34.0% (out of 209; 95% CI: 27.7%-40.9%); self-harmed by overdosing/poisoning, 66.4% (95% CI: 60.5%-71.9%) vs. 76.1%; (95% CI: 69.6%-81.6%), and had a history of diagnosis of any mental health condition, 63.2% (95% CI: 57.2%-68.8%) vs. 74.2% (95% CI: 67.6%-79.8%). Although not statistically evident, more individuals self-harmed by hanging/strangulation in the DC group (4.3% vs. <2.0%). Differences in distributions of sex and age group were not significantly different between DC and BC groups in the ED and enhanced dataset (Suppl. Table 9 in ESM 3).

10-year Self-harm and mortality outcomes

From the enhanced dataset, we identified 157 (56.1% out of 280) in the DC group and 123 (58.9% out of 209) individuals in the BC group who self-harmed during the 10-year follow-up, with no statistical evidence for group differences (unadjusted OR: 0.9, 95% CI: 0.6-1.3, p = 0.580; Bayes factors: 0.1-0.3, evidence in favor of independence between self-harm and group, Fig. 1 and Suppl. Table 7-8 in ESM 3). All-cause mortality was higher in the DC than the BC group (unadjusted OR = 1.9, 95% CI: 1.0-3.6, p = 0.047; Bayes factors: 3.9-11.5, moderate/strong evidence in favor of dependence between all-cause mortality and group). More individuals in the DC group, died by natural causes. Mean age of death, mortality by unnatural causes and suicide were similar between groups. Results from Firth logistic regressions show statistically higher mortality for males in the DC group compared to other three groups (Suppl. Table 10-11 in ESM 3). Older age group was also statistically associated with higher mortality.

VIFs for all independent variables in all corresponding adjusted regressions for this study ranged between 1.0 and 2.2, which were lower than the adopted threshold of three. This suggests that multicollinearity was not an issue for all our adjusted models.

Discussion

For the first time to our knowledge, this study compared characteristics of individuals who self-harmed during a suicide cluster with those who self-harmed one year before and followed them for up to 10 years for self-harm and mortality outcomes. While our observation of higher number individuals who self-harmed during the cluster might reflect an actual increase, it could also be due to the heightened awareness and thus change in behavior of recording self-harm from clinicians at the time of the cluster in comparison to the situation where self-harm were under-reported or poorly recorded out of the period of the cluster. We found an increase in the number of individuals who self-harmed during the cluster but with less related hospitalisation, which may reflect self-harm severity, methods used or clinical practice during a cluster with increased demand. It may also reflect policy/practice to reduce public concerns. SAD scores, and histories of self-harm was similar between groups. There was some evidence of greater use of hanging as a method for self-harm during the cluster, consistent with methods widely reported in the media at the time (Marchant, Turner, et al., 2020). Individuals who self-harmed during the suicide cluster were similarly likely to those from the non-cluster to repeat self-harm over the 10-years follow-up. Males who self-harmed during the cluster had higher long-term all-cause mortality risks. Since these findings were not predicted a priori and require replication and the contributing factors remain unclear, further investigations on long-term outcomes are warranted (Haw et al., 2013).

Strengths and limitations

This unique study compared individuals who self-harmed during a suicide cluster with non-cluster self-harm cases and evaluating long-term self-harm and mortality outcomes by linking clinical assessment to routinely collected data. The high data coverage in the SAIL databank facilitated comparisons of individual characteristics and increased sample size by identifying individuals using diagnostic codes for self-harm. However, small sample size is still a huge issue in this study. We used both frequentist and Bayesian approaches to test our hypotheses and results were in tight agreement between approaches. We collected ED admission data from a single hospital only as this hospital is the only district general hospital providing secondary care services covering the relevant LAA. We included individuals based on geographical proximity only and not on psychological or social proximity, which are important factors in clustering of suicides (Hawton et al., 2020); data and measures for these two dimensions are required in future research. We excluded a small number of individuals who self-harmed during both pre-cluster (BC) and cluster periods (DC) to ensure tenability of data stratification and statistical analyses. While the corresponding proportions to the whole

datasets were small (<3.5%), such exclusion may still introduce bias particularly for the BC group, which may be less likely to experience outcomes in the 10-year follow-up. As for other research using routinely collected data, we are likely to underestimate self-harm for those who do not contact health services or have their conditions misclassified.

Implications for policy and practice

Our findings can inform intervention strategies to prepare for, identify, and respond to suicide clusters (Public Health England, 2019). Increased self-harm risk during a cluster is not confined to those with pre-existing mental health diagnoses and long-term outcomes of those who self-harm are broader. We highlight a potential need for long-term monitoring and intervention in those who self-harm during suicide clusters. While it is crucial to identify and provide timely interventions/support to vulnerable individuals following suicide clusters, attention should also be paid to the general health and wellbeing of the whole community, particularly for males following a cluster.

Authors biographies

Sze Chim Lee, PhD, is a senior research data scientist in Medicine, Health and Life Science at Swansea University Medical School. His research uses administrative data and surveys to study a range of biological, psychosocial, and environmental circumstances that may be associated with mental health issues, suicide, and self-harm.

Olivier Y. Rouquette, PhD, is researcher and data scientist in the Population Psychiatry, Suicide and Informatics (PPSI) team at Swansea University Medical School, working in prof. Ann John's team. Olivier's research encompasses mental health and wellbeing of children and young people using routinely collected population linked data.

Keith Hawton is professor of Psychiatry and Director of Centre for Suicide Research at the University of Oxford. Professor Hawton has a particular interest in epidemiology and clinical management of self-harm, suicide and self-harm in adolescents, media influences on self-harm and evaluation of suicide prevention initiatives.

Louise Cleobury, PhD, is Senior Lecturer in Health Data Science and Programme Director for Population Health and Medical Sciences at Swansea University Medical School. Louise's areas of interest are in Clinical, Applied, and Health Psychology. Louise Cleobury has over 15 years' experience in multidisciplinary applied health research across settings.

Sarah Spencer is retired from a successful career in NHS including: Emergency Medicine consultant, Head of Postgraduate Training in Emergency Medicine, Clinical Director Acute &

Emergency Services, Deputy Medical Director, Locality Group Director (Primary, Community, Secondary Acute and Mental Health Services).

Keith Lloyd is professor of psychiatry at Swansea University Medical School and a clinical academic specialising in psychiatry. Keith’s research interests are in epidemiology, suicide, and the use of routine health data in mental health research. He is pro-vice chancellor for medicine, health, and life science at Swansea University.

David Gunnell, FMedSci, is Emeritus Professor of Epidemiology at the University of Bristol, UK. He is a public health physician and epidemiologist with a longstanding research interest in the etiology and prevention of suicide and in improving population mental health.

Jonathan Scourfield is Professor of Social Work and Deputy Director of CASCADE, the Children's Social Care Research and Development Centre at Cardiff University. His research includes child and family services, working with men, social work education, research capacity-building, the social context of suicide and self-harm, and identity and religion in children.

Ann john is Professor in Public Health and Psychiatry at the Swansea University Medical School. She chairs the National Advisory Group to Welsh Government on the prevention of suicide and self-harm. Her research targets suicide, self-harm prevention and mental health with an emphasis on translating research into policy and practice.

Electronic Supplementary Material

– ESM 1. RECORD checklist (RECORD_Checklist.docx).

– ESM 2. Suppl. Methods (Suppl_Methods.docx).

The document shows additional descriptions of methodology and statistical analysis.

– ESM 3. Suppl. Tables 1-11 and Suppl. Fig. 1 (Suppl_Tables_Figures.docx).

The document shows additional tables and figures not shown in the main text.

References

Carr, M. J., Ashcroft, D. M., Kontopantelis, E., While, D., Awenat, Y., Cooper, J., Chew-Graham, C., Kapur, N., & Webb, R. T. (2017). Premature Death Among Primary Care Patients With a History of Self-Harm. *The Annals of Family Medicine*, *15*(3), 246–254. https://doi.org/10.1370/AFM.2054

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38. https://doi.org/10.1093/biomet/80.1.27

Ford, D. V, Jones, K. H., Verplancke, J., Lyons, R. A., John, G., Brown, G., Brooks, C. J., Thompson, S., Bodger, O., Couch, T., & Leake, K. (2009). The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Services Research*, *9*(1), 157. https://doi.org/10.1186/1472-6963-9-157

Haw, C., Hawton, K., Niedzwiedz, C., & Platt, S. (2013). Suicide Clusters: A Review of Risk Factors and Mechanisms. *Suicide and Life-Threatening Behavior*, *43*(1), 97–108. https://doi.org/10.1111/j.1943-278X.2012.00130.x

Hawton, K., Hill, N. T. M., Gould, M., John, A., Lascelles, K., & Robinson, J. (2020). Clustering of suicides in children and adolescents. *The Lancet Child & Adolescent Health*, *4*(1), 58–67. https://doi.org/10.1016/S2352-4642(19)30335-9

Hawton, K., Lascelles, K., & Ferrey, A. (2015). *Identifying and responding to suicide clusters and contagion: A practical resource*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/839621/PHE_Suicide_Cluster_Guide.pdf

Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, *21*(16), 2409–2419. https://doi.org/10.1002/sim.1047

John, A., DelPozo-Banos, M., Gunnell, D., Dennis, M., Scourfield, J., Ford, D. V, Kapur, N., & Lloyd, K. (2020). Contacts with primary and secondary healthcare prior to suicide: case–control whole-population-based study using person-level linked routine data in Wales, UK, 2000–2017. *The British Journal of Psychiatry*, *217*(6), 717–724. https://doi.org/DOI: 10.1192/bjp.2020.137

John, A., Hawton, K., Gunnell, D., Lloyd, K., Scourfield, J., Jones, P. A., Luce, A., Marchant, A., Platt, S., Price, S., & Dennis, M. S. (2016). Newspaper Reporting on a Cluster of Suicides in the UK. *Crisis*, *38*(1), 17–25. https://doi.org/10.1027/0227-5910/a000410

John, A., Marchant, A., Hawton, K., Gunnell, D., Cleobury, L., Thomson, S., Spencer, S., Dennis, M., Lloyd, K., & Scourfield, J. (2022). Understanding suicide clusters through exploring self-harm: Semi-structured interviews with individuals presenting with near-fatal self-harm during a suicide cluster. *Social Science & Medicine*, *292*, 114566. https://doi.org/https://doi.org/10.1016/j.socscimed.2021.114566

John, A., McGregor, J., Jones, I., Lee, S. C., Walters, J. T. R., Owen, M. J., O'Donovan, M., DelPozo-Banos, M., Berridge, D., & Lloyd, K. (2018). Premature mortality among people with severe mental illness — New evidence from linked primary care data. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2018.04.009

Joiner, T. E. (1999). The Clustering and Contagion of Suicide. *Current Directions in Psychological Science*, *8*(3), 89–92. https://doi.org/10.1111/1467-8721.00021

Jones, P., Gunnell, D., Platt, S., Scourfield, J., Lloyd, K., Huxley, P., John, A., Kamran, B.,

Wells, C., & Dennis, M. (2013). Identifying Probable Suicide Clusters in Wales Using National Mortality Data. *PLoS ONE*, *8*(8), e71713. https://doi.org/10.1371/journal.pone.0071713

Lyons, R. a, Jones, K. H., John, G., Brooks, C. J., Verplancke, J.-P., Ford, D. V, Brown, G., & Leake, K. (2009). The SAIL databank: linking multiple health and social care datasets. *BMC Medical Informatics and Decision Making*, *9*, 3. https://doi.org/10.1186/1472-6947-9-3

Marchant, A., Brown, M., Scourfield, J., Hawton, K., Cleobury, L., Dennis, M., Lloyd, K., McGregor, J., & John, A. (2020). A Content Analysis and Comparison of Two Peaks of Newspaper Reporting During a Suicide Cluster to Examine Implications for Imitation, Suggestion, and Prevention. *Crisis*, *41*(5), 398–406. https://doi.org/10.1027/0227-5910/a000655

Marchant, A., Turner, S., Balbuena, L., Peters, E., Williams, D., Lloyd, K., Lyons, R., & John, A. (2020a). Self-harm presentation across healthcare settings by sex in young people: an e-cohort study using routinely collected linked healthcare data in Wales, UK. *Archives of Disease in Childhood*, *105*(4), 347 LP – 354. https://doi.org/10.1136/archdischild-2019-317248

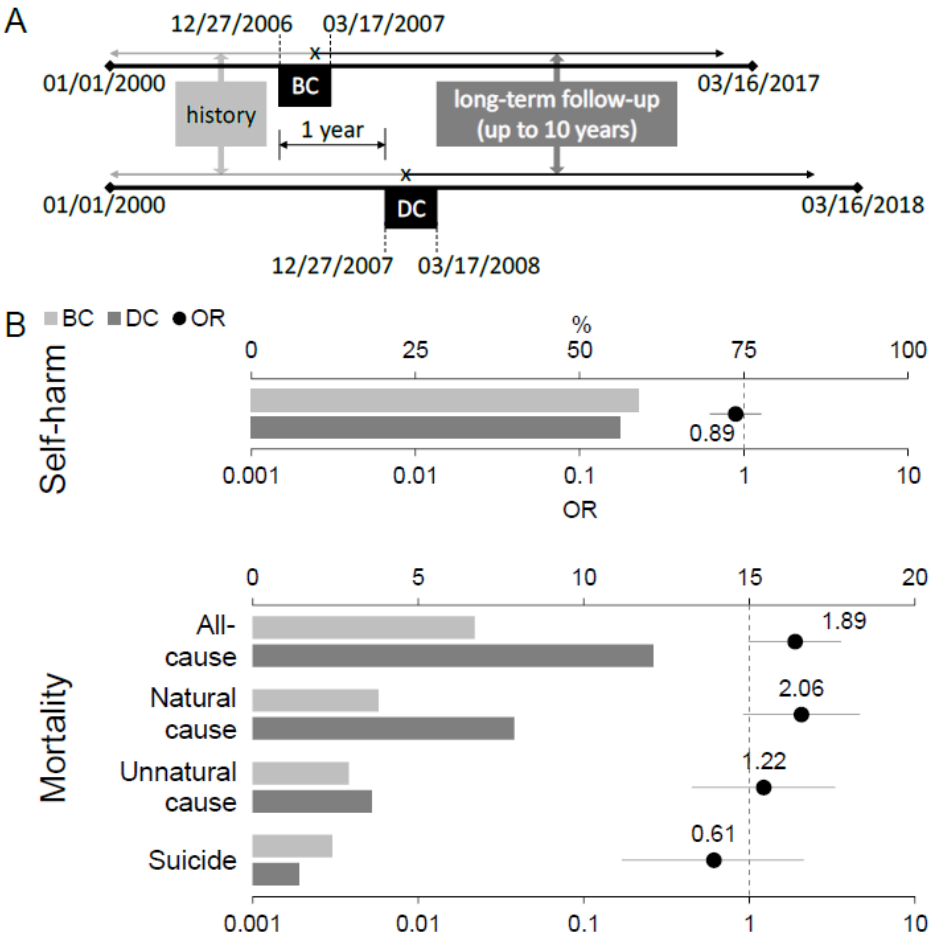
Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. Sage.

Office for National Statistics. (2019). *Suicides in the UK: 2018 registrations*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/suicidesintheunitedkingdom/2018registrations

Patterson, W. M., Dohn, H. H., Bird, J., & Patterson, G. A. (1983). Evaluation of suicidal patients: The SAD PERSONS scale. *Psychosomatics*, *24*(4), 343–349. https://doi.org/10.1016/S0033-3182(83)73213-5

Public Health England. (2019). *Identifying and responding to suicide clusters and contagion: A practice resource.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/839621/PHE_Suicide_Cluster_Guide.pdf

Robinson, J., Pirkis, J. & O'Connor, R.C. (2016). Suicide Clusters. In R.C. O'Connor and J. Pirkis (Eds.), *The International Handbook of Suicide Prevention* (2nd ed., pp. 758-774). Wiley Blackwell. https://doi.org/10.1002/9781118903223.ch43



398

399 Fig. 1. (A) Schematic diagram of observation period of this study. DC: Self-harm
400 ascertainment period during to the suicide cluster (December 27, 2007-March 17, 2008); BC:
401 Self-harm ascertainment period one year before the suicide cluster (December 27, 2006-
402 March 17, 2007); X: index self-harm event during ascertainment period. (B) Comparison of
403 self-harm and mortality outcomes during a 10-year follow-up. Odds ratios (ORs) are
404 analysed by univariable Firth regression. Error Bars: 95% CIs.

405 **The RECORD statement – checklist of items, extended from the STROBE statement, that should be**
406 **reported in observational studies using routinely collected health data.**

407

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Title and abstract
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			Introduction
Objectives	3	State specific objectives, including any prespecified hypotheses			Introduction
Methods					
Study Design	4	Present key elements of study design early in the paper			Methods (Study design)
Setting	5	Describe the setting, locations, and relevant dates,			Methods (settings), Fig. 1 and

		including periods of recruitment, exposure, follow-up, and data collection			Suppl. Methods in ESM 2
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	Methods, Suppl. Methods in ESM 2, Fig. 1 and Suppl. Fig. 1 in ESM 3
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be	Methods, Suppl. Methods in ESM 2

				reported, an explanation should be provided.	
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			Methods, Suppl. Methods in ESM 2
Bias	9	Describe any efforts to address potential sources of bias			Methods, Suppl. Methods in ESM 2
Study size	10	Explain how the study size was arrived at			Methods, Suppl. Methods in ESM 2
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Methods, Suppl. Methods in ESM 2
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of			Methods, Suppl. Methods in ESM 2

		cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Methods, Suppl. Methods in ESM 2
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Methods, Suppl. Methods in ESM 2
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage.		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Results and Suppl. Fig. 1 in ESM 3

		(c) Consider use of a flow diagram			
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			Results, Suppl. Table 1-4 and Suppl. Fig. 1 in ESM 3
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			Results, Suppl. Table 5 and Fig. 1
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized			Results, Suppl. Table 5 and Fig. 1

		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			Results and Suppl. Table 6-7
Discussion					
Key results	18	Summarise key results with reference to study objectives			Discussion
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Discussion
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			Discussion
Generalisability	21	Discuss the generalisability (external validity) of the study results			Discussion
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if			Title page

		applicable, for the original study on which the present article is based			
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Title page

408

409 *Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von
410 Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using
411 Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

412

413 *Checklist is protected under Creative Commons Attribution (CC BY) license.

414

Suppl. Methods

A sample ED mental health assessment form

Bwrdd Iechyd Prifysgol Abertawe Bro Morgannwg University Health Board

ED Mental Health Assessment Form

ED Dr _____ **Date** _____ **Time** _____ : _____

ED No _____
Surname _____
First Name _____
DOB _____
Address _____

Telephone: _____
Mobile: _____

GP Name _____
Practice _____
Fax No _____

NOK: _____
Relationship: _____
Contact No: _____

History

Previous history ☐ **Known to Ψ services** ☐ **None** ☐ **Unknown**
Details *(method/date of previous self harm, details of mental health input, substance use etc)*

Social circumstances

☐ Single ☐ Separated/divorced ☐ Partnered/married ☐ Widowed

☐ Lives alone ☐ With partner/spouse ☐ With parents
☐ Lone parent ☐ With friends/relatives ☐ Homeless

☐ Dependent adult(s) ☐ Dependent child(ren) ☐ No dependents

Notes

Medication, drug/substance/alcohol use

Allergies _____

rev 30 July 2014

417

Patient's full name _____ DOB _____

SAD PERSONS SCORE

Sex (M = 1, F = 0)	☐	Rational thinking (organic illness or psychosis = 2)	☐
Age (<19 or >45 =1)	☐	Separated/divorced/widowed (yes = 1)	☐
Depression/Hopelessness (present = 2)	☐	Organised/serious attempt (yes = 2)	☐
Previous attempts or in-pt care (yes = 1)	☐	No social supports (No support = 1)	☐
Excessive alcohol/drug use (yes = 1)	☐	Stated future intent (Yes = 2)	☐

TOTAL ☐

(0 – 5 : consider discharge, 6 – 8: ref psych, 9 – 14 : likely needs admission... **THIS IS ONLY A GUIDE!**

Mental State Examination *(Appearance/Behaviour, Mood (subj/obj), Speech (vol/rate/flow), Thought (content/process), Perception (hallucinations), Orientation, Cognitive function, Insight)*_____

Does the patient have capacity to give informed consent to any intervention that may be necessary?
YES/NO
If YES and the patient is declining treatment, senior or psychiatric review is MANDATORY before the patient is allowed to leave the hospital.

Impression

Plan (Completion mandatory). **All patients age ≥ 65 MUST be referred for formal MH assessment**)

☐ Admit Physicians ☐ Admit ITU ☐ Admit Surgical specialty

☐ Refer Psychiatry
Time of referral: _____

☐ Home, notes faxed to Crisis Resolution Team on ext 2673, **give patient information sheet so they can contact CRHTT**

☐ Home, Ψ F/U already arranged
(ie patient already known to MH services)

☐ Home, notes faxed to GP for GP F/U (this should be the default option for most patients)

418

419

420

Other Covariates

We used sex, age (as group: 0-14, 15-34, 35-54, and 55 years or above), marital status (single, separated/divorced/widowed, partnered/married, and unknown), household status (lives alone, with lone parent, with parents, with partner/spouse, with friends/relatives, and Others & unknown), area deprivation, and urban/rural indicator as sociodemographic variables. Area deprivation was categorized according to quintiles of Welsh Index of Multiple Deprivation (WIMD) 2011 score for all lower-layer super-output areas (LSOAs) in Wales (Welsh Government, 2011), with the first quintile (Q1) represents the least and the fifth (Q5) the most deprived areas. The urban/rural indicator for England and Wales was used to categorize urban and rural LSOAs (Barham & Begum, 2006). LSOAs with unknown WIMD quintile or urban/rural indicator were grouped as "unknown LSOA" category. Please note that marital and household status were not used for the enhanced dataset due to the data unavailability in SAIL databank.

Apart from demographic variables, we included the following variables from the SAIL datasets (primary and secondary care), as shown in a previous study on premature mortality following self-harm (Carr et al., 2017).

Physical comorbidity: We used the Charlson comorbidity index (CCI) to measure individuals' physical comorbidity (Charlson et al., 1987). The CCI is based on 17 binary scores for the presence of any of the 17 physical illnesses. We used both Read (for primary care dataset) and International Classification of Diseases version 10 (ICD-10, for secondary care dataset) codes to identify each of these illnesses based on previous studies (Bottle & Aylin, 2011; Khan et al., 2010). The unweighted CCI, i.e., summing the binary score for an individual was calculated and categorized into two groups: CCI =0 and CCI ≥1.

Previous self-harm events as a binary variable: Self-harm events were extracted using previously used Read and ICD-10 code lists (Carr et al., 2017; Marchant et al., 2020; Thomas et al., 2013).

Any mental health diagnoses as a binary variable: Mental health diagnoses were extracted according to the definitions used in a previous study (Ann John et al., 2020). All Read codes within the category of mental disorders (E…. and all associated subcodes) and ICD-10 codes within the category of mental and behavioural disorders (F00-F99) were used for this variable.

Common mental disorders (CMDs) as a binary variable: We used previously used code lists (John et al., 2015; John, Marchant, et al., 2016; John, McGregor, et al., 2016) to identify individuals with CMDs, including mainly depression and anxiety, from both SAIL datasets.

Severe mental illness (SMI) as a binary variable: We adopted the definition and the code lists of SMI used in previous studies (Economou et al., 2012; Ford et al., 2009; John et al., 2018; Lloyd et al., 2015). This included schizophrenia, schizotypal, delusional, and schizoaffective disorders, bipolar disorder, and other psychotic disorders.

Alcohol and drug misuse: We used the previously defined code lists for alcohol (Carr et al., 2017; John et al., 2020; McKenzie et al., 2010; Quan et al., 2005) and drug misuse (John et al., 2020; Quan et al., 2005; Thompson et al., 2004). Alcohol and drug misuse were separately represented by two binary variables.

Prescription of psychotropic and opiates drug medications: Prescription of drug medications could be extracted from the primary care dataset only. For psychotropic medications including antidepressants, anxiolytics, hypnotics, and antipsychotics, we used the code definition from others (Dennis et al., 2017; John et al., 2015; John et al., 2020). Code lists adopted by (John et al., 2020) were used for opiates medication. We used two separate binary variables for psychotropic and opiates drug medications.

All described variables were time-fixed and age, marital status, household status, area deprivation, and urbanicity were measured as at the date of the index self-harm event. Other variables were measured from January 01, 2000 to the date before the date of index self-harm event (defined as history period, see Suppl. Fig. 1 in ESM 3).

Analysis and statistical methods

Linked data in SAIL were interrogated using structured query language (SQL DB2). Statistical analyses were conducted using Stata 17 and R. Level of statistical significance was set at $p = 0.05$. We compared individuals' characteristics and outcomes between DC and BC groups. Number of individuals who self-harmed and the methods of self-harm used were compared between DC and BC group during their two respective ascertainment periods as well as during the 10-year follow-up period. All descriptive statistics were summarized as person counts and percentages or group means for continuous variables with 95% confidence intervals (CIs). CIs for proportions were estimated by Wilson score with continuity correction (Newcombe, 1998).

Due to the issue of small sample size, we used both frequentist and Bayesian approaches (Jamil et al., 2017; Oliveira et al., 2018), including Fisher's exact tests, likelihood ratio tests, and Bayes factors to examine independence of variables for contingency tables. While Fisher's exact and likelihood ratio tests are classical hypothesis tests for independence that associated with *p*-values, Bayes factors directly estimate the weights of evidence over two competing hypotheses, i.e., dependence vs. independence of variables in contingency tables. Bayes factor reflects the degree of shift of beliefs about the relative odds between the two hypotheses (Jeffreys, 1961). We reported Bayes factor as a ratio of the conditional probabilities associated with the alternative (dependence) to those with the null hypothesis (independence) given the observed data. Thus, Bayes factor >1 and <1 respectively represent evidence in favor of the alternative and null hypothesis and a Bayes factor of unity indicates no evidence towards any of the hypothesis. All Bayes factors for contingency tables were calculated using the 'BayesFactor' package in R (Morey et al., 2022) and we reported range of bayes factors based on the four available data sampling plans (Poisson, joint multinomial, independent multinomial, and hypergeometric) and used uninformative priors with concentration parameter of one.

Differences in means for continuous variables between groups were assessed by independent sample *t* test accompany with the corresponding Bayes factors of the *t* statistics. Similarly, Bayes factor >1 and <1 respectively provide evidence for and the presence (alternative hypothesis) and absence (null hypothesis) of mean differences between groups. All Bayes factors for *t* test were calculated using the 'BayesFactor' package in R (Morey et al., 2022) and we reported the range of bayes factors based on the Cauchy priors with scale parameters of $\sqrt{2}/2$, 1, and $\sqrt{2}$.

We interpreted all Bayes Factors in this study using the previously reported guidelines: 1-3 as providing anecdotal, 3-10 as moderate, 10-30 as strong, 30-100 as very strong, and >100 as extreme evidence for the alternative hypothesis (Jamil et al., 2017).

Effect modification of stratified cross-tabulation by sex and age was tested by the homogeneity of odds ratios (ORs) based on the Breslow-Day test adjusted by (Tarone, 1985). We also reported the effect sizes (as ratio of ORs) of the sex-by-age group interaction term from the Firth logistic regression of the probability of being in BC or DC group, with sex, age group, and the interaction term as predictors.

For the enhanced dataset, we built multivariable regression models for self-harm and mortality outcomes during the 10-year follow-up when significant difference between DC and BC group exists in the descriptive statistics. We stratified the DC and BC groups further by

sex (BC-male, BC -female, DC-male and DC-female) and performed the adjusted analysis with two models. Model 1 adjusted for age group (≤34 vs. >34 years), area deprivation (WIMD quintile), and urban/rural indicator. Model 2 included all variables in Model 1 and further adjusted for the CCI, history of self-harm, any mental health diagnoses, CMD, SMI, alcohol misuse, drug misuse, prescription of psychotropic, and prescription of opiate medications. We performed Firth logistic regression (Firth, 1993) to circumvent biased estimates from conventional maximum likelihood estimation due to small sample size and separation issues (Heinze & Schemper, 2002). To evaluate the differences in all-cause mortality among the four sex-stratified DC and BC groups, we computed multiple pairwise comparisons following regression modelling and reported the Wald chi-square statistics and the corresponding unadjusted and Holm-adjusted p-values. While we reported estimates from the Firth logistic regression in the main text, results from conventional logistic regression were also shown in relevant supplementary tables for reference. We conducted diagnostic checks on multicollinearity by calculating the variance inflation factors (VIFs) of all independent variables in all adjusted models. We used the commonly adopted VIF threshold of three to determine if multicollinearity is an issue for each model (Miles & Shevlin, 2001).

References for Suppl. Methods

Barham, C., & Begum, N. (2006). The new urban/rural indicator in the Labour Force Survey. *Labour Market Trends*, *114*(12), 409–418.

Bottle, A., & Aylin, P. (2011). Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices. *Journal of Clinical Epidemiology*, *64*(12), 1426–1433. https://doi.org/10.1016/j.jclinepi.2011.04.004

Carr, M. J., Ashcroft, D. M., Kontopantelis, E., While, D., Awenat, Y., Cooper, J., Chew-Graham, C., Kapur, N., & Webb, R. T. (2017). Premature Death Among Primary Care Patients With a History of Self-Harm. *The Annals of Family Medicine*, *15*(3), 246–254. https://doi.org/10.1370/AFM.2054

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases*, *40*(5), 373–383. https://doi.org/10.1016/0021-9681(87)90171-8

Dennis, M., Shine, L., John, A., Marchant, A., McGregor, J., Lyons, R. A., & Brophy, S. (2017). Risk of Adverse Outcomes for Older People with Dementia Prescribed Antipsychotic Medication: A Population Based e-Cohort Study. *Neurology and Therapy*, *6*(1), 57–77. https://doi.org/10.1007/s40120-016-0060-6

Economou, A., Grey, M., McGregor, J., Craddock, N., Lyons, R. A., Owen, M. J., Price, V., Thomson, S., Walters, J. T. R., & Lloyd, K. (2012). The Health Informatics Cohort Enhancement project (HICE): using routinely collected primary care data to identify people with a lifetime diagnosis of psychotic disorder. *BMC Research Notes*, *5*, e95. https://doi.org/10.1186/1756-0500-5-95

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38. https://doi.org/10.1093/biomet/80.1.27

Ford, D. V, Jones, K. H., Verplancke, J., Lyons, R. A., John, G., Brown, G., Brooks, C. J., Thompson, S., Bodger, O., Couch, T., & Leake, K. (2009). The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Services Research*, *9*(1), 157. https://doi.org/10.1186/1472-6963-9-157

Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, *21*(16), 2409–2419. https://doi.org/10.1002/sim.1047

Jamil, T., Ly, A., Morey, R. D., Love, J., Marsman, M., & Wagenmakers, E.-J. (2017). Default "Gunel and Dickey" Bayes factors for contingency tables. *Behavior Research Methods*, *49*(2), 638–652. https://doi.org/10.3758/s13428-016-0739-8

Jeffreys, H. (1961). *Theory of Probability: Oxford Classic Texts in the Physical Sciences* (3rd ed.). Oxford University Press.

John, A., Marchant, A. L., Fone, D. L., Mcgregor, J. I., Dennis, M. S., Tan, J. O. A., & Lloyd, K. (2016). Recent trends in primary-care antidepressant prescribing to children and young people: An e-cohort study. *Psychological Medicine*, *46*(16), 3315–3327. https://doi.org/10.1017/S0033291716002099

John, A., Marchant, A. L., McGregor, J. ., Tan, J. O. A., Hutchings, H. A., Kovess, V., Choppin, S., Macleod, J., Dennis, M. S., & Lloyd, K. (2015). Recent trends in the incidence of anxiety and prescription of anxiolytics and hypnotics in children and young

people: An e-cohort study. *Journal of Affective Disorders*, *183*, 134–141. https://doi.org/10.1016/J.JAD.2015.05.002

John, A., McGregor, J., Fone, D., Dunstan, F., Cornish, R., Lyons, R. A., & Lloyd, K. R. (2016). Case-finding for common mental disorders of anxiety and depression in primary care: An external validation of routinely collected data. *BMC Medical Informatics and Decision Making*, *16*(1), 35. https://doi.org/10.1186/s12911-016-0274-7

John, Ann., McGregor, J., Jones, I., Lee, S. C., Walters, J. T. R., Owen, M. J., O'Donovan, M., DelPozo-Banos, M., Berridge, D., & Lloyd, K. (2018). Premature mortality among people with severe mental illness — New evidence from linked primary care data. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2018.04.009

John, Ann, DelPozo-Banos, M., Gunnell, D., Dennis, M., Scourfield, J., Ford, D. V, Kapur, N., & Lloyd, K. (2020). Contacts with primary and secondary healthcare prior to suicide: case–control whole-population-based study using person-level linked routine data in Wales, UK, 2000–2017. *The British Journal of Psychiatry*, *217*(6), 717–724. https://doi.org/DOI: 10.1192/bjp.2020.137

Khan, N. F., Perera, R., Harper, S., & Rose, P. W. (2010). Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. *BMC Family Practice*, *11*, 1. https://doi.org/10.1186/1471-2296-11-1

Lloyd, K., McGregor, J., John, A., Craddock, N., Walters, J. T., Linden, D., Jones, I., Bentall, R., Lyons, R. A., Ford, D. V., & Owen, M. J. (2015). A national population-based e-cohort of people with psychosis (PsyCymru) linking prospectively ascertained phenotypically rich and genetic data to routinely collected records: Overview, recruitment and linkage. *Schizophrenia Research*, *166*(1–3), 131–136. https://doi.org/10.1016/j.schres.2015.05.036

Marchant, A., Turner, S., Balbuena, L., Peters, E., Williams, D., Lloyd, K., Lyons, R., & John, A. (2020). Self-harm presentation across healthcare settings by sex in young people: an e-cohort study using routinely collected linked healthcare data in Wales, UK. *Archives of Disease in Childhood*, *105*(4), 347 LP – 354. https://doi.org/10.1136/archdischild-2019-317248

McKenzie, K., Harrison, J. E., & McClure, R. J. (2010). Identification of alcohol involvement in injury-related hospitalisations using routine data compared to medical record review. *Australian and New Zealand Journal of Public Health*, *34*(2), 146–152. https://doi.org/10.1111/j.1753-6405.2010.00499.x

Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers.* Sage.

Morey, R. D., Rouder, J. N., Urbanek, S., Forner, K., & Ly, A. (2022). *BayesFactor: Computation of Bayes Factors for Common Designs*. https://cran.r-project.org/web/packages/BayesFactor/index.html

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proposition: comparison of seven methods. *Statistics in Medicine*, *17*(8), 857–872. https://doi.org/10.1002/(SICI)1097-0258(19980430)17

Oliveira, N. L., Pereira, C. A. de B., Diniz, M. A., & Polpo, A. (2018). A discussion on significance indices for contingency tables under small sample sizes. *PLOS ONE*, *13*(8), e0199102. https://doi.org/10.1371/journal.pone.0199102

620 Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J.-C., Saunders, L. D.,
621 Beck, C. A., Feasby, T. E., & Ghali, W. A. (2005). Coding Algorithms for Defining
622 Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Medical Care*, *43*(11),
623 1130–1139. www.jstor.org/stable/3768193

624 Tarone, R. E. (1985). On heterogeneity tests based on efficient scores. *Biometrika*, *72*(1),
625 91–95. https://doi.org/10.1093/biomet/72.1.91

626 Thomas, K. H., Davies, N., Metcalfe, C., Windmeijer, F., Martin, R. M., & Gunnell, D. (2013).
627 Validation of suicide and self-harm records in the Clinical Practice Research Datalink.
628 *British Journal of Clinical Pharmacology*, *76*(1), 145–157.
629 https://doi.org/10.1111/bcp.12059

630 Thompson, A., Shaw, M., Harrison, G., Ho, D., Gunnell, D., & Verne, J. (2004). Patterns of
631 hospital admission for adult psychiatric illness in England: analysis of Hospital Episode
632 Statistics data. *British Journal of Psychiatry*, *185*(4), 334–341. https://doi.org/DOI:
633 10.1192/bjp.185.4.334

634 Welsh Government. (2011). *Welsh Index of Multiple Deprivation 2011 Summary Report*.
635 http://webarchive.nationalarchives.gov.uk/20120404081001/http://wales.gov.uk/topics/s
636 tatistics/publications/wimd11summary/?lang=en

637

638

639