BMJ Open

Suicide Information Database: A protocol for a population based, routinely collected data linkage case control study to explore risks and patterns of healthcare contact prior to suicide to identify opportunities for intervention.

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-006780
Article Type:	Protocol
Date Submitted by the Author:	29-Sep-2014
Complete List of Authors:	John, Ann; Swansea University, College of Medicine Dennis, Mick; Swansea University, College of Medicine Kosnes, Liv; Swansea University, College of Medicine Gunnell, David; Bristol University, School of Social and Community Medicine Scourfield, Jonathan; Cardiff University, School of Social Sciences Ford, David; Swansea University, Health Information Research Unit Lloyd, Keith; Swansea University, College of Medicine
Primary Subject Heading:	Public health
Secondary Subject Heading:	Epidemiology, General practice / Family practice, Health informatics, Mental health, Public health
Keywords:	Suicide & self-harm < PSYCHIATRY, PUBLIC HEALTH, EPIDEMIOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

Suicide Information Database- Cymru: A protocol for a population based, routinely collected data linkage study to explore risks and patterns of healthcare contact prior to suicide to identify opportunities for intervention.

John A[1*], Dennis M[1], Kosnes L[1], Gunnell D[2], Scourfield J[3], Ford DV[1], Lloyd K[1]

[1]College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park, Swansea, SA2 8PP, United Kingdom.

[2]School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, United Kingdom.

[3]School of Social Sciences, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff, CF10 3WT, United Kingdom.

Corresponding Author: Ann John[1] a.john@swansea.ac.uk; Phone: +44(0)1792 602568; College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park, Swansea, SA2 8PP, United Kingdom.

Michael Dennis[1] m.s.dennis@swansea.ac.uk, Liv Kosnes[1] l.kosnes@swansea.ac.uk, David Gunnell[2] d.j.gunnell@bristol.ac.uk , Jonathan Scourfield[3] Scourfield@cardiff.ac.uk, David V. Ford[1] d.v.ford@swansea.ac.uk, Keith Lloyd[1] k.r.lloyd@swansea.ac.uk

Key words: Suicide, case-control, database, routine data linkage.

Word Count: 4054

Abstract

Introduction

Prevention of suicide is a global public health challenge extending beyond mental health services. Linking routinely collected health and social care system data records, for the same individual across different services and over time has enormous potential in suicide research. Most previous research linking suicide mortality data with routinely collected electronic health records involves only one or two domains of health care provision such as psychiatric in-patient care. This protocol paper describes the development of a population based, routinely collected data linkage study: the Suicide Information Database (SID-Cymru). SID-Cymru aims to contribute to the information available on people who complete suicide.

Methods and Analysis

SID-Cymru will facilitate a series of electronic case-control studies based in the Secure Anonymised Information Linkage (SAIL) Databank. We have identified 2664 cases of suicide in Wales between 2003 and 2011 from routinely collected mortality data using International Classification of Diseases, Tenth Revision, codes X60-84 (intentional self harm) and Y10-34 (undetermined intent). Each case will be matched by age and sex to at least five controls who are alive at the time of death of the case. Records will be collated and linked from routinely collected health and social data in Wales for each individual. Conditional logistic regression will be applied to produce crude and confounder (including general practice, socioeconomic status) adjusted odds ratios.

Ethics and Dissemination

The SAIL Databank has the required ethical permissions in place to analyse anonymised data. Ethical approval has been granted by the Information Governance Review Panel (IGRP). The

improved understanding of the prior health, nature of previous contacts with services, and wider social circumstances of those who complete suicide will assist in prevention policy, service organisation and delivery. SID-Cymru is funded through the National Institute for Social Care and Health Research, Welsh Government (RFS-12-14).

Article Summary

Article Focus:

- Case definitions of probable and possible suicides
- Identifying risk factors and patterns of service use in those who complete suicide
- Designing a study

Key messages:

- SID-Cymru will facilitate a series of case- control studies of those who complete suicide
- Recency of primary care, in- and out-patient hospital and emergency department contacts including attendance for self-harm, primary care diagnosis of depression, levels of treatment with anti-depressants and trends in such treatment over time, rural and urban geography, educational attainment, contacts for the elderly and levels of physical illness will be investigated
- Results will be adjusted for a number of confounders

Strengths and Limitations of this study:

- Based on the general population
- Includes all deaths through suicide, not just those known to mental health services

- Based on routinely collected health and social care data linkage
- Linkage across multiple domains. Most suicide data linkage studies are limited to only one or two domains of health care provision.
- Prescriptions not actual use

Introduction

Every year approximately 800,000 people die by suicide worldwide, one to two in every 100 deaths. Prevention of suicide is a global public health challenge. Collaborative working across government departments, with a public health approach extending beyond mental health service care is essential[1]. Global patterns and national trends in the incidence of suicide and its key risk factors change over time. The therapeutic and preventive challenges of understanding and responding to these changes are considerable.

Vast amounts of personal data are routinely collected on a daily basis by health and social care systems around the world to support clinical management and patient care. Linking these data records, for the same individuals across different services and over time, offers a powerful, population-wide resource. Such integrated datasets have been used to study a range of health issues to identify risk and protective factors and to examine outcomes. The secondary use of these data has enormous potential in suicide research. Improved consideration of the prior health, wider social circumstances and points of access to services of all individuals who complete suicide can be achieved.[2-4]

Studies from the Nordic countries, have demonstrated the usefulness of data-linkage across register based studies in suicide research.[5-10] Others[11-13] have demonstrated that collating and linking sets of routinely collected whole population based data, such as GP records, out-patient data, in-patient activity, with mortality data enable more detailed analysis of risk factors for those people completing suicide. Most previous research linking suicide mortality data with routinely collected electronic health records involves only one or two domains of health care provision such as psychiatric in-patient care.

In the United Kingdom various systems exist to examine suicide deaths. The National Confidential Inquiry into Suicide and Homicide (NCISH) by people with Mental Illness focuses on suicide cases that were in contact with mental health services (approximately 25% of total*)* in the year prior to their deaths.[3] However this provides limited information on issues of suicide in the general population and may hamper the effectiveness of wider preventive efforts.[1] In a recent report NCISH[14] examined aspects of primary healthcare prior to all suicides in England between 2002-2011; however no linkage was made with data from other service providers such as emergency departments.

Scotland has recently established enhanced data collection in relation to suicide, however, further development is needed in order for 'ScotSID' to be able to examine healthcare pathways and contact with more than one health service;[15] England does not currently have a dedicated repository for suicide data though studies have made use of the Clinical Practice Research Datalink (CPRD; www.cprd.com), which represents approximately 8.5% of the UK population from 600 General Practices (GP's) in England.[16] CPRD can be linked with data from the National Health Service (NHS) Hospital Episode Statistics (HES; http://www.hscic.gov.uk) and mortality data from the Office of National Statistics (ONS; http://www.ons.gov.uk) but has limited emergency department data.

SID-Cymru will access and link information on prior health, nature of previous contacts with services, and wider social circumstances of all those who complete suicide (known and unknown to mental health services) within the population of Wales via anonymised routinely collected electronic data held in healthcare and social datasets from the Secure Anonymised Information Linkage (SAIL) Databank (www.saildatabank.com).[17 18] This brings together, links and anonymises the widest possible range of person-based data currently available in

the UK. The SAIL Databank was originally set up by the Health Information Research Unit (HIRU) at the College of Medicine at Swansea University.

SID-Cymru is part of the research programme related to the Health e-Research Collaboration UK (HeRC UK), led by the Medical Research Council (MRC) and based in the Centre for the Improvement of Population Health through e-Records Research (CIPHER). CIPHER is a UK Clinical Research Collaboration (UKCRC) Public Health Research Centre of Excellence set within the College of Medicine at Swansea University. SID-Cymru will provide timely robust data to inform the future strategic direction of and the first step in designing and evaluating effective interventions to prevent suicide. It will use the International Classification of Disease, Tenth Revision,[19] (ICD-10) definitions and instructions for classifying causes of death which will allow for comparisons with other countries and thus support research of relevance globally. An emphasis will be given to issues where there are opportunities for intervention or where electronic data linkage confers an advantage to its investigation.

Aim

The main aim is to establish a Suicide Information Database- Cymru as a population based resource for studying factors and service contacts associated with all suicide deaths through routinely collected data linkage case control studies.

Objectives:

Phase 1:

- To identify, via the SAIL Databank, all those with ICD-10 codes for probable and possible suicide in Wales 2003-2011, and matched controls.

- To explore and address the methodological issues relating to the development of this database of completed suicides and the linkage of data across different datasets and settings.

Phase 2:

- To investigate risk factors and trends for suicide including: primary care diagnosis of depression; levels of treatment with anti-depressants and trends in such treatment over time; rural and urban geography; educational attainment; levels of physical illness.
- To investigate settings and pathways of care where people are in contact with services in the year leading up to their suicide across the whole population and in specific groups such as the elderly.

Methods and Analyses

Design

SID-Cymru will facilitate a series of electronic population based, routinely collected data linkage case-control studies on completed suicide in Wales between 2003 and 2011. Wales has a population of 3.1 million[20] and is part of the United Kingdom. There are approximately 32,000 deaths of Welsh nationals registered each year of which around 300 (approximately 1%) are registered as suicides.[21]

Data Source

SID-Cymru will be assembled within the SAIL Databank,[17] a growing resource that already holds over a billion anonymised records from 13 databases which can be anonymously linked at the individual record level.[18] The SAIL Databank has been previously used for linkage of routine data.[22-26]

Within the SAIL Databank a split-file approach to anonymisation is used to overcome issues of confidentiality and disclosure in health-related data warehousing.[18] Demographic data are sent to a partner organisation, NHS Wales Information Service (NWIS), where all identifiable information is removed; clinical data are sent directly to HIRU, where, for each dataset within the SAIL Databank, an individual is assigned an encrypted Anonymised Linking Field (ALF). The ALF is used to link anonymised individuals across datasets, thus supporting the opportunity to conduct longitudinal analyses of an individual's journey through multiple health, education and social datasets.[17] Additionally Residential Anonymised Linking Fields (RALFs) have been created for all residences in Wales and enables linkage of anonymised household and environment data with the health records of individual residents without the identity of the residences or residents being known to researchers.[26]

The primary study base will be the Welsh Demographic Service[27] (WDS) hosted within the SAIL Databank. The WDS is a core dataset available within the SAIL Databank and part of a set of services to manage administrative information (demographic data) for NHS patients in Wales. The WDS was introduced early in 2009 replacing a similar service known as the NHS Wales Administrative Register (NHS AR). The WDS data is collected from GP practices via the Exeter System and more than five million individuals are currently present in the WDS dataset within the SAIL Databank. The WDS is a register of all individuals that have at some point in time been registered with a Welsh General Practitioner (GP) or required some form of NHS health care provision in Wales. The electronic collation of WDS / NHS AR data originated in 1960 and is updated and maintained by NWIS,[27] ensuring that address changes (within and out of Wales) and death notices are included in the register. The original (non-anonymised) version of the NHS AR has been used in the HIRU Matching Algorithm for Consistent Results in Anonymised Linkage (MACRAL), making the WDS / NHS AR the

master list for all Welsh residents and using probabilistic matching to find the associated NHS numbers that are then encrypted into ALF's.

Deaths in Wales should be registered within five days of the date of death. However legislation in Wales means that when a coroner's inquest takes place, the death cannot be registered until the inquest is complete. Since the Office for National Statistics (ONS), the national agency where all deaths are collated, has no conclusive information about the death until it is registered, there is a delay between the date the death occurred and when the death is added to the annual ONS mortality dataset. The ONS found that less than 41% of deaths going to inquest were registered within 3 months, though 96% were registered within 1 year.[28] Thus precise information on the annual incidence of suicide may be delayed by up to two years. Information collected at death registration is recorded on the Registration Online system by registrars. Most of the information is normally supplied by the informant (usually a close relative of the deceased) while the cause of death (COD) is usually obtained from the Medical Certificate of Cause of Death (MCCD) completed by a medical practitioner when the death is certified using ICD-10 coding or the coroner if there is an inquest.[19] Notably, a death is not officially registered within the Annual District Deaths Extract (ADDE) until the COD has been finalised, and thus the year of death and the year of registration may not concur. The *primary dataset* used to construct SID-Cymru is the ADDE from the ONS. The ADDE is inclusive of Welsh residents who died outside of Wales, and holds information about COD derived from death certificates on all deaths in Wales.

Definition of Suicide for Cases

The true number of suicides is difficult to determine because a coroner's conclusion of suicide must be 'beyond a reasonable doubt' i.e. that the death was intentionally self inflicted and in some areas coroners have increasingly (since 2001) reported narrative conclusions

rather than reporting it as suicide.[29 30] Previously when insufficient information was recorded by the coroner ONS coders used to record the death as an accident which inevitably led to some suicides being classified as accidents or misadventure. The ONS has recently issued guidance on this issue following a coding practice review.[31] Current ONS practice includes deaths where intention is 'undetermined whether accidentally or purposefully inflicted'; thus deaths where there may be no intention to take life, such as in relation to injury or poisoning, are included in suicide figures by ONS. Currently there is no access to coroners' narrative verdicts within the SAIL Databank as a possible method for review of case inclusion.

There is evidence to suggest that a high proportion of deaths from poisoning and hanging that receive accidental verdicts, are found, when subjected to clinical review, to be suicides.[32] Such possible deaths through suicide will be included in SID-Cymru as an opportunity for further separate and combined analysis; thus the additional ICD-10 codes relating to 'accidental poisoning with prescribed drugs' (X40-X41, X43-X49) and 'accidental hanging' (W75-W76) may be used along with 'sequelae of external causes of morbidity and mortality' (Y87, Y87.2, Y89, Y89.9).[18] Thus with SID-Cymru we aim to establish a Suicide Information Database that can be a resource for studying factors associated with *all* probable, as well as, possible deaths through suicide; by including accidental hanging/strangulation and accidental poisoning, excluding, narcotics and psychodysleptics (i.e., possible suicides), along with the ONS traditional method of defining suicides, which includes suicides and deaths of undetermined intent (probable suicides).[2]

Identification of Cases and Controls

SID-Cymru cases encompass suicides ('intentional self-harm' [ICD-10: X60-84]) and probable deaths through suicide ('undetermined intent' [ICD-10: Y10-34, excl.Y33.9]) recorded from the MCCD and presented within ADDE as the underlying COD. Those coded

Y33.9/U50.9 (pending verdicts) are excluded since a large proportion of these are subsequently found to be homicides. Probable suicide defined above can be supplemented with possible deaths through suicide if required. Cases of probable suicide will be identified and extracted by use of the ICD-10 codes[19] defined above and depicted in Table 1.

ONS figures include those over 15 years of age only, due to the possibility that deaths in younger children coded as undetermined events may be caused by unverifiable accidents, neglect or abuse.[33] SID-Cymru will allow for the analysis of suicides and probable suicides for the 10 to 14 year age band. Official ONS mortality statistics are produced based on the number of deaths registered in a particular calendar year, rather than the number of deaths that occurred in that year. This means their figures include some deaths that occurred in years prior to the reference year (approximately 4%). As SID-Cymru will link and review data in the period leading up to death it is important to ensure the match/end date reflects the correct time period (specific to each individual case), i.e. date of death not date of registration, to afford an accurate perspective on utilisation of resources and help seeking behaviours. Consequently the actual date of death (ADOD) will be used in the matching criteria to establish a data review 'end date' for controls, rather than the registered date of death (RDOD) referred to in ONS reports. Mortality data within the SAIL Databank is only available from 2003, thus the earliest case inclusion relates to ADOD's from the 1st January 2003. To minimise the underestimation of cases identified (i.e., due to delays in COD confirmation and registration of DOD) the case inclusion for SID-Cymru includes cases aged 10 years and over at ADOD, where the ADOD took place between 01/01/2003 and 31/12/2011.

Table 1. ICD-10 codes used to identify and extract cases with death through probable suicide for SID-Cymru.

Short name	**Description**	**ICD10 codes**
Poisoning	Poisoning by and exposure to one or more types of poisons	X60-X69, Y10-Y19
Hanging & suffocation	Hanging, strangulation and suffocation	X70, Y20
Drowning	Drowning and submersion	X71, Y21
Guns	Firearm discharge	X72-X74, Y22-Y24
Sharp object	Sharp object	X78, Y28
Jumping from a high place	Jumping from a high place	X80, Y30
Jumping or lying before moving object	Jumping or lying before moving object	X81, Y31
Explosives & Fire	Explosive material, smoke, fire, flames and hot objects	X75-X77, Y25-Y27
Other or unspecified events	Intentional self-harm by other/un- specified means; Other/un- specified events undetermined intent	X83-X84, Y33-Y34 (excl.Y33.9)
Crash by motor vehicle	Crashing of motor vehicle intentional and undetermined intent	X82, Y32
Blunt Object	Blunt Object	X79, Y29

Table 2 presents the number of registered deaths through suicide per year and actual deaths per year between 2003 and 2011 as reported by the ONS and as identified for SID-Cymru within the SAIL Databank.

Table 2. Probable Suicide[1] Deaths for Wales 2003 – 2011 identified within the SAIL Databank for SID-Cymru.

Count by Year[2]	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**TOTAL**
Probable Suicides ADOD[3] within SID[4]	361	303	287	278	279	282	272	292	304	**2664**
Probable Suicides RDOD[5] within SID	274	330	270	300	288	266	257	287	340	**2612**
ONS Probable Suicides[6] RDOD	332	330	269	300	289	266	258	288	341	**2673**

[1]Probable suicides = ICD-10 codes for Intentional [X60-84], Undetermined [Y10-34 excl.Y33.9] self-harm; [2]Values are number of individuals unless otherwise indicated [3]ADOD = Year death occurred; [4]Cases identified for SID-Cymru includes those aged 10 years and over; [5]RDOD = Year death was registered; [6]Cases reported by the ONS include those aged 15 years and over.

Matched Controls will be identified within the WDS as live individuals matched on age (to the nearest year) and gender who were registered within the WDS for at least one year prior to matched case's ADOD. The controls will be required to be alive at the time of the matched case's ADOD (i.e., the match/end date). The use of live controls limits the introduction of bias relating to deaths, particularly in the younger age groups, of those undertaking risky behaviours resulting in premature death (e.g. substance use, accidents) that may be associated with unrecognised suicidal behaviours or known risk factors. To add to the power of the study we aim to identify at least five controls to every one case.[34]

Routine Data Sources

For SID-Cymru the data collected on identified cases and controls, via ADDE and the WDS respectively, will be linked to other routinely collected datasets, retrospectively allowing a review of each individual's pathway through the various services. Linkage with, for example, GP system data provides varying information about patients going back several years, including previous diagnosis, presenting symptoms and previous medications prescribed. This dataset can be used to review contacts with the GP and, consequently, infer the development / diagnosis of any new medical conditions including depression and self harm prior to suicide. Linkage with in-patient data will allow a review of hospital contacts and Emergency Department datasets will give information on crisis contacts. This will provide comprehensive insights into help seeking behaviours and management across settings. Data sets currently accessible via the SAIL Databank which will be linked to SID-Cymru are presented in Table 3. There are several other datasets currently under negotiation for inclusion within the SAIL Databank to which, in the future, SID-Cymru could link. These include Department of Work and Pensions (DWP) employment and incapacity status, Looked After Children, fostering, Substance Misuse Services, Sexually Transmitted Infections and Police Data.

Table 3. Datasets available within the SAIL Databank for linkage with SID-Cymru cases and controls.

Dataset	Data Description	Data Provider	Data Type	Geographical Coverage	Available from
Annual District Death Extract (ADDE)	Register of all deaths relating to Welsh residents, including those that died out of Wales.	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW) on behalf of ONS	Registry data	Wales	2003
Welsh Demographic Service (WDS)	Administrative information about individuals in Wales that use NHS services; such as address and practice registration history. It replaced the NHS Wales Administrative Register (NHSAR) in 2009.	NHS Wales' Informatics Service (NWIS)/ HSW	Administrative (mandatory) data return	Wales	1990
Area Based Deprivation Scores	Lower Super Output Area - Townsend Deprivation Index and Welsh Index of Multiple Deprivation Scores.	NHS Wales' Informatics Service (NWIS)/ HSW	Administrative data	Wales	2001 / 2005
Education Attainment (DCEL)	Educational attainment at key stages (pre 16 year old). Data collated from school census, the national pupil survey and educational attainment databases by the Welsh Government.	Department for Children, Education, Lifelong Learning and Skills (DCELLS)	Administrative / survey data.	Wales	2005
Primary Care GP dataset	Each GP practice uses a clinical information system to maintain an electronic health record for each of their patients; capturing the signs, symptoms, test results, diagnoses, prescribed treatment, referrals for specialist treatment and social aspects relating to the patients home environment.	Welsh General Practices that have signed up to the SAIL Databank	Clinical system data	Wales	2000
Patient Episode Database for Wales (PEDW)	NHS Wales - hospital admissions (Inpatients and day cases) dataset comprising of attendance and clinical information for all hospital admissions: includes diagnoses and operations performed.	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW)	Administrative (mandatory) data return	Wales	1997
Outpatient Dataset (OPD)	Attendance information for all NHS Wales hospital outpatient appointments.	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW)	Administrative (mandatory) data return	Wales	2004
Emergency Department Data Set (EDDS)	Administrative and clinical information for all NHS Wales Accident and Emergency department attendances. Includes the All Wales Injury Surveillance Systems (AWISS) dataset	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW)	Administrative (mandatory) data return	Wales	2009

Measurements

Data Variables/Characteristics

Data to be included in SID-Cymru will be extracted from the SAIL Databank and will include basic demographics; educational data; ADOD and RDOD; numbers and percentages for deprivation; proportions known to different healthcare settings in the period prior to death; inclusive medical history, i.e., primary care contact and diagnosis (by Read Codes), information about hospital / psychiatric admission and diagnosis (by ICD-10 Codes), and nature of service contact e.g. for self-harm, substance misuse. Therefore, the routinely collected data held by the NHS and other public bodies supplying the existing SAIL Databank will maximise the narrative of a death through suicide while being less resource intensive than psychological autopsies. The initial variables to be extracted and linked across data sets are described in Table 4.

Planned Analyses

The primary objective is to establish SID-Cymru as a resource for future analysis. It is important to maximise the utility of the resource, and some general principles of analysis have been determined.

Descriptive Epidemiology

Phase 1:

- Identification and description of Cases: Number of deaths registered with ONS and available within the SAIL Databank, with relevant ICD-10 codes defined as suicides, between 2003-2011.

Table 4. Characteristics to be identified and collected through data linkage for SID-Cymru Cases and Controls.

Data	Characteristics
Demographics	Sex
	Date of Birth
	Marital status*
	Employment*
	Education*
Mortality	Date of Death,
	Age at DOD (in years)
	Registered Date of Death
	Cause of Death (ICD-10 code)
Deprivation	LSOA
	WIMD
	Townsend
	Rural-Urban Indicators
Medical History	GP register*
	GP diagnosis (Read code)*
	GP events*
	Hospital attendance
	Hospital diagnosis
	Hospital events
	Outpatient attendance
	Outpatient diagnosis*
	Outpatient events
	ED attendance*
	ED diagnosis*
	ED events*
	Psychiatric Admission*

* Where data are available; there may be limitations to the number of cases with this information.

- Identification and description of Matched Controls.
- Basic demographics.
- Table of delay (days) in registering suicides and undetermined deaths.

Phase 2:

- Proportions known to different healthcare settings: number and percentages, with main diagnosis, that had a general hospital admission; emergency department contact for self-harm and other indications; psychiatric admission and primary care contact in the year prior to probable suicide.
- Numbers and percentages for deprivation, employment status, educational achievement and medical history (e.g., chronic pain, terminal illness, medication, previous self harm, and substance misuse) will also be sought.
- Number of cases with missing data across data sets for variables of interest will be noted.

Area based measures of socio-economic deprivation

Deprivation will be measured at Lower Super Output Area (LSOA) level using the Welsh Index of Multiple Deprivation[35] (WIMD) and Townsend Index Score.[36] All suicides and matched controls will be assigned to a lower layer super output area (LSOA). There are 1909 LSOAs in Wales with an average population of 1500 people (range: 1000-3000).[37] Linkage to WIMD and Townsend Index information is available in the SAIL Databank. These will be ranked for deprivation, divided into quintiles and standardised rates calculated.

Odds ratios for the described exposures in the case control study

A case-control study utilising SID-Cymru will be population based and so the relative risk of suicide will be estimated by conditional logistic regression model with SPSS (version 20). Crude odds ratios will be adjusted for general practice and/or LSOA by matching cases and controls. Unadjusted estimates, confounder-adjusted estimates and their precision (e.g., 95% confidence interval) will be produced. Interactions between variables will be assessed with the log likelihood ratio test based on results from the adjusted analysis. The population attributable risk will be calculated[38] on the basis of adjusted relative risks from the full analysis and the distribution of exposures in the cases.

We will also report information on the completeness of linkage with each dataset.

Ethics and Dissemination

Ethics

A large amount of preliminary work on anonymisation methodologies was undertaken to create the SAIL Databank system,[17- 18] and the SAIL Databank has the required ethical permissions and processes in place to analyse anonymised data. It operates within a robust series of guidelines in line with the Caldicott principles and the National Information Governance Board for Health and Social Care.[18] Ethical approval has been granted for SID-Cymru from HIRU's Information Governance Review Panel (IGRP), an independent body consisting of a range of government, regulatory and professional agencies. In compliance with the IGRP rulings and the Data Protection Act 1998[39] individual level data and personal identifier linkage codes will not be removed from the SAIL Databank and all analyses will be carried out within the SAIL Databank gateway at Swansea University, a secure access point to data within the SAIL Databank. The key points of the MRC/Wellcome Trust data sharing policy will be followed.[40 41]

This paper describes the protocol for the development of a Suicide Information Database (SID-Cymru), and the research opportunities available from an electronic case-control study of suicides within a whole population. SID-Cymru will have the ability to link suicide cases anonymously to primary and secondary health information along with other social care data, allowing us to review each case's journey through these datasets. The establishment of SID-Cymru and exploration of the linkage methodologies will improve our understanding of those who complete suicide (particularly those not known to mental health services) and will be used to inform service planning and policy decision making and implementation. It will help identify key opportunities and settings for prevention of this tragic event. By so doing, SID-Cymru will join other international databases of suicide research and provide a platform for further investigation and data linkages.

In order for SID-Cymru to become a functional resource it is important to be aware of the limits of health data available; though widely used in research, and offering a broad range of information about treatment and associated conditions, there are issues relating to determining the quality of patient records, the completeness of data available and any conclusions that may be drawn from them, perhaps particularly concerning primary care records.[42] That is, working with routinely collected data presented in its "raw" format requires a certain level of database analysis skills where duplicates, missing and erroneous entries are common occurrences. Whilst some such administrative / system-based recording issues are easy to identify and account for in individual datasets, it is not always apparent what is correct and what is erroneous at the combined level. Indeed, this problem is confounded when linkage of data reveals conflicting information causing routine data to appear inaccessible and attempts at linkage discouraged. Thus, a secondary aim for SID-Cymru is to share the skills developed as part of establishing a suicide database, which can

aid colleagues who may lack such analytical expertise and foster greater multi-disciplinary collaborations and advance suicide research.

The UK has a strong presence in form of a wide range of publications and expertise relating to suicide research. Successful and dedicated Suicide Research Centres exist in Bristol, Manchester and Oxford,[43-45] and Scotland recently commenced work on a Scottish Suicide Information Database (‘ScotSID’).[46] These centres of excellence report broadly about suicide, though for logistical reasons, often a regional focus is retained, relating to project funding, analytical and interventional work; for that reason the establishment of a Suicide Information Database in Wales will build on and enhance existing UK suicide resources and infrastructure since the level of linkage available within the SAIL Databank is unique in this field. Unique opportunities within the SAIL Databank include linkage across primary, secondary and emergency department data and with education data.

Implications and Significance

Valuable opportunities exist for a wide range of epidemiological and clinical studies on suicide in Wales and SID-Cymru has the potential to become an important resource in facilitating such research which will be of relevance internationally. In addition to the records that have already been included for linkage with SID-Cymru it is expected that over time relevant information from other data sources (e.g. the DWP) will be linked to SID-Cymru to provide a wider range of information on issues such as individuals’ circumstances, the nature of their deaths, and their contact with extended services. Additionally, linkage can be made with non-routinely collected datasets such as that held by NCISH. Specific hypotheses that will be explored include: recency of primary care, hospital and emergency department contacts including attendance for self-harm, primary care diagnosis of depression, levels of treatment with anti-depressants and trends in such treatment over time, rural and urban geography, contacts for the elderly and levels of physical illness.

Findings of current and projected public health importance will be assessed and presented to support policy makers, commissioners and providers of health and social care in Wales. Non-identifiable information from this project will be made available to researchers in Wales, the UK and international collaborators. The initial project focus will be on identification of cases and controls, data linkage opportunities and methodological issues relating to the establishment of a Suicide Information Database and routine data linkage (Phase 1), before commencing data extraction and analysis (Phase 2). The proposed data collation and linkage of primary, secondary and emergency department health information together with educational data are currently unique in the United Kingdom. Due to existing routine data linkages within the SAIL Databank[17] it is possible to develop a central repository for information relating to suicide in whole population that will be of relevance internationally. This paper describes the design and development of a Suicide Information Database in its infancy.

Authors' contributions

AJ, the principal investigator, was responsible for the conceptualization of SID-Cymru. AJ, KL, MD, DG, JS for the design of SID-Cymru; AJ and LK for its on-going operationalisation. AJ and LK drafted the manuscript. All authors read and approved the final manuscript.

Funding Statement

This study was funded by a grant from the National Institute for Social Care and Health Research, Welsh Government, grant number RFS-12-14. DG is a National Institute for Health Research (England) Senior Investigator.

Competing interests

The authors declare that they have no competing interests.

Ethics approval

Ethical approval has been granted by the Health Information Research Unit Information Governance Review Panel at the College of Medicine at Swansea University.

Data sharing statement

It will be possible to access the data after the publication of the results. Researchers interested in collaborations or further information are invited to contact AJ at a.john@swansea.ac.uk.

References

1. World Health Organisation. 66th World Health Assembly. Draft comprehensive mental health action plan 2013-2020. A66/10 Rev.1, WHO 16 May 2013. http://apps.who.int/gb/e/e_wha66.html Accessed September 2014.
2. Linsley KR, Schapira K, Kelly TP. Open verdict v. suicide—importance to research. *Br J Psychiatry* 2001, **178**:465-468.
3. Appleby L, Kapur N, Shaw J, *et al*. Annual Report: England, Northern Ireland, Scotland and Wales. Manchester: University of Manchester 2013. http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/AnnualReport2013_UK.pdf Accessed July 2014.

4. Andersen UA, Andersen M, Rosholm JU, *et al*. Contacts to the health care system prior to suicide: a comprehensive analysis using registers for general and psychiatric hospital admissions, contacts to general practitioners and practising specialists and drug prescriptions. *Acta Psychiatr Scand* 2000, **102**:126-34.
5. Qin P, Agerbo E, Mortensen PB. Suicide risk in relation to socioeconomic, demographic, psychiatric, and familial factors: a national register–based study of all suicides in Denmark, 1981–1997. *Am J Psychiatry* 2003, **160**:765-772.
6. Gunnell D, Löfving S, Gustafsson JE, *et al*. School performance and risk of suicide in early adulthood: follow-up of two national cohorts of Swedish schoolchildren. *Journal of affective disorders*, 2011, **131**:104-112.
7. Qin P, Hawton K, Mortensen PB, *et al*. Combined effects of physical illness and comorbid psychiatric disorder on risk of suicide in a national population study. *Br J Psychiatry* 2014, **204**:430-435.
8. Christiansen E, Larsen KJ, Agerbo E, *et al*. Incidence and risk factors for suicide attempts in a general population of young people: a Danish register-based study. *Aust N Zealand J Psychiatry* 2012, **47**: 259-270.
9. Björkenstam C, Johansson LA, Nordström P, *et al*. Suicide or undetermined intent? A register-based study of signs of misclassification. *Population health metrics*, 2014, **12**:11.
10. Hjorthøj CR, Madsen T, Agerbo E, *et al*. Risk of suicide according to level of psychiatric treatment: a nationwide nested case–control study. *Social psychiatry and psychiatric epidemiology*, 2014; 1-9.

11. Lawrence DM, Holman CD, Jablensky AV, *et al*. Suicide rates in psychiatric in-patients: an application of record linkage to mental health research**.** *Aust NZ J Public Health* 1999, **23**:468–470.

12. Lawrence D, D'Arcy C, Holman J, *et al.* Increasing rates of suicide in Western Australian psychiatric patients: a record linkage study. *Acta Psychiatr Scand* 2001, **104**:443-451.

13. Morrison KB, Laing L. *Adults' use of health services in the year before death by suicide in Alberta.* Statistics Canada; 2011.

14. Appleby L, Kapur N, Shaw J, *et al*. Suicide in primary care in England: 2002-2011. National Confidential Inquiry into Suicide and Homicide by People with Mental Illness (NCISH). Manchester: University of Manchester 2014. http://www.bbmh.manchester.ac.uk/cmhr/research/centreforsuicideprevention/nci/reports/SuicideinPrimaryCare2014.pdf Accessed August 2014.

15. ISD Scotland: The Scottish Suicide Information Database Report 2014. http://www.isdscotland.org/Health-Topics/Public-Health/Publications/index.asp?ID=1232 Accessed August 2014.

16. Webb RT, Kontopantelis E, Doran T, *et al*. Suicide risk in primary care patients with major physical diseases: a case-control study. *Arch Gen Psychiatry* 2012, **69**:256-264.

17. Lyons RA, Jones KH, John G, *et al*. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Making* 2009, **9**:3.

18. Ford DV, Jones KH, Verplancke JP, *et al*. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009, **9**:1.

19. World Health Organization: *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. WHO: Geneva; 1993.

20. Office of National Statistics. 2011 Census Results for Wales. ONS 16th July 2012. http://wales.gov.uk/docs/statistics/2012/120716-census-2011-first-results-wales-en.pdf Accessed September 2014.

21. Office of National Statistics. Statistical bulletin: Suicide rates in the United Kingdom, 2006 to 2010. ONS 26th January 2012. http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2010/stb-statistical-bulletin.html Accessed August 2014.

22. McGregor JI, Brooks CJ, Chalasani P, *et al*. The Health Informatics Trial Enhancement Project (HITE): Using routinely collected primary care data to identify potential participants for a depression trial. *Trials* 2010, **11**:39.

23. Brooks CJ, Stephens JW, Price DE, *et al*. Use of a patient linked data warehouse to facilitate diabetes trial recruitment from primary care. *Primary Care Diabetes* 2009, **3**:245-248

24. Hill RA, Brophy S, Brunt H, *et al*. Protocol of the baseline assessment for the Environments for Healthy Living (EHL) Wales cohort study. *BMC Public Health* 2010, **10**:150

25. Fone DL, Dunstan F, White J, *et al*. Cohort Profile: The Caerphilly Health and Social Needs Electronic Cohort Study (E-CATALyST). *International Journal of Epidemiology*, 2013, **42**:1620-1628.

26. Rodgers SE, Heaven M, Lacey A, et al. Cohort profile: the housing regeneration and health study. *Int J Epidemiol* 2012, dys200.

27. SAIL Databank: Core Data Sets. Swansea University: Swansea http://www.saildatabank.com/data-dictionary/sail-datasets Accessed September 2014.

28. Brock A, Baker A, Griffiths C, *et al*. Suicide trends and geographical variations in the United Kingdom, 1991–2004. *Health Statistics Quarterly* 2006, **31**:6-22.

29. Carroll R, Hawton K, Kapur N, *et al*. Impact of the growing use of narrative verdicts by coroners on geographic variations in suicide: analysis of coroners' inquest data. *Journal of Public Health* 2011, **10**:1-7.

30. Gunnell D, Hawton, K, Kapur N. Coroners' verdicts and suicide statistics in England and Wales. *BMJ* 2011, **343**:d6030.

31. Hill C, Cook, L. Narrative verdicts and their impact on mortality statistics in England and Wales. *Health Statistics Quarterly* 2011, **49**:81-100.

32. Gunnell D, Bennewith O, Simkin S, *et al*. Time trends in coroners' use of different verdicts for possible suicides and their impact on officially reported incidence of suicide in England: 1990-2005. *Psychol Med* 2013, **43**:1415-1422.

33. Office of National Statistics. Statistical bulletin: Suicide rates in the United Kingdom, 2011. ONS 22 January 2013. http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2011/stb-suicide-bulletin.html Accessed August 2014.

34. Grimes DA, Schultz KF. Compared to what? Finding controls in a case control study. *Lancet* 2005, **365**: 1429–33.

35. Office for National Statistics. Welsh index of Multiple Deprivation 2011. Summary report, Welsh Government; 2011.

36. Townsend P, Phillimore P, Beattie A. *Health and Deprivation: Inequality and the North*. Routledge: London; 1988.

37. Office for National Statistics. Super Output Areas 2011. http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html Accessed August 2014.

38. Bruzzi P, Green SB, Byar DP, *et al*. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol* 1985, **122**:904-914.

39. Great Britain: Data Protection Act 1998. London: HMSO. http://www.legislation.gov.uk/ukpga/1998/29/contents Accessed August 2014.

40. MRC: Policy and Guidance on Sharing of Research Data from Population and Patient Studies. MRC 23rd November 2011. http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC008302 Accessed August 2014.

41. Wellcome Trust: Data sharing policy. Wellcome Trust August 2010. http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm Accessed August 2014.

42. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ* 2003, **326**:1070.

43. University of Bristol, Suicide Research Centre. http://www.bristol.ac.uk/psychiatry/research/suicideresearch.html Accessed August 2014.

44. The University of Manchester Centre for Suicide Prevention. http://www.bbmh.manchester.ac.uk/cmhr/research/centreforsuicideprevention/ Accessed August 2014.

45. University of Oxford, Centre for Suicide Research. http://cebmh.warne.ox.ac.uk/csr/ Accessed August 2014.

46. ISD Scotland. Scottish Suicide Information Database (ScotSID). ISD Scotland 2011. http://data.gov.uk/dataset/scottish_suicide_information_database Accessed August 2014.

STROBE Statement—checklist of items that should be included in reports of observational studies-protocol limits some of these

	Item No	Recommendation	Line
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	33
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	34-41
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	73-131
Objectives	3	State specific objectives, including any prespecified hypotheses	136-150
Methods			
Study design	4	Present key elements of study design early in the paper	152-7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	158-300
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	209-282
		(*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	237-282
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Table 3 , table 4, 283-349
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	Table 3 , table 4, 283-349
Bias	9	Describe any efforts to address potential sources of bias	350-360
Study size	10	Explain how the study size was arrived at	245-282, table 2
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	314-359:protocol
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	314-359:protocol
		(*b*) Describe any methods used to examine subgroups and interactions	314-359:protocol
		(*c*) Explain how missing data were addressed	314-359:protocol
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed	236-282

	Case-control study—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy
	(*e*) Describe any sensitivity analyses

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Table 2, protocol
		(b) Give reasons for non-participation at each stage	Will
		(c) Consider use of a flow diagram	Will
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Will
		(b) Indicate number of participants with missing data for each variable of interest	Will
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
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	Will
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Will
		(*b*) Report category boundaries when continuous variables were categorized	Will
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Will
Discussion			
Key results	18	Summarise key results with reference to study objectives	Will
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	Will
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Will
Generalisability	21	Discuss the generalisability (external validity) of the study results	Will
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	447

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Suicide Information Database: A protocol for a population-based, routinely collected data linkage study to explore risks and patterns of healthcare contact prior to suicide to identify opportunities for intervention.

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-006780.R1
Article Type:	Protocol
Date Submitted by the Author:	14-Oct-2014
Complete List of Authors:	John, Ann; Swansea University, College of Medicine Dennis, Mick; Swansea University, College of Medicine Kosnes, Liv; Swansea University, College of Medicine Gunnell, David; Bristol University, School of Social and Community Medicine Scourfield, Jonathan; Cardiff University, School of Social Sciences Ford, David; Swansea University, Health Information Research Unit Lloyd, Keith; Swansea University, College of Medicine
Primary Subject Heading:	Public health
Secondary Subject Heading:	Epidemiology, General practice / Family practice, Health informatics, Mental health, Public health
Keywords:	Suicide & self-harm < PSYCHIATRY, PUBLIC HEALTH, EPIDEMIOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

Suicide Information Database- Cymru: A protocol for a population-based, routinely collected data linkage study to explore risks and patterns of healthcare contact prior to suicide to identify opportunities for intervention.

John A[1*], Dennis M[1,] Kosnes L[1], Gunnell D[2], Scourfield J[3], Ford DV[1], Lloyd K[1]

[1]College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park, Swansea, SA2 8PP, United Kingdom.

[2]School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, United Kingdom.

[3]School of Social Sciences, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff, CF10 3WT, United Kingdom.

Corresponding Author: Ann John[1] a.john@swansea.ac.uk; Phone: +44(0)1792 602568; College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park, Swansea, SA2 8PP, United Kingdom.

Michael Dennis[1] m.s.dennis@swansea.ac.uk, Liv Kosnes[1] l.kosnes@swansea.ac.uk, David Gunnell[2] d.j.gunnell@bristol.ac.uk , Jonathan Scourfield[3] Scourfield@cardiff.ac.uk, David V. Ford[1] d.v.ford@swansea.ac.uk, Keith Lloyd[1] k.r.lloyd@swansea.ac.uk

Key words: Suicide, case-control, database, routine data linkage.

Word Count: 4054

Abstract

Introduction

Prevention of suicide is a global public health challenge extending beyond mental health services. Linking routinely collected health and social care system data records, for the same individual across different services and over time has enormous potential in suicide research. Most previous research linking suicide mortality data with routinely collected electronic health records involves only one or two domains of health care provision such as psychiatric in-patient care. This protocol paper describes the development of a population based, routinely collected data linkage study: the Suicide Information Database (SID-Cymru). SID-Cymru aims to contribute to the information available on people who complete suicide.

Methods and Analysis

SID-Cymru will facilitate a series of electronic case-control studies based in the Secure Anonymised Information Linkage (SAIL) Databank. We have identified 2664 cases of suicide in Wales between 2003 and 2011 from routinely collected mortality data using International Classification of Diseases, Tenth Revision, codes X60-84 (intentional self harm) and Y10-34 (undetermined intent). Each case will be matched by age and sex to at least five controls. Records will be collated and linked from routinely collected health and social data in Wales for each individual. Conditional logistic regression will be applied to produce crude and confounder (including general practice, socioeconomic status) adjusted odds ratios.

Ethics and Dissemination

The SAIL Databank has the required ethical permissions in place to analyse anonymised data. Ethical approval has been granted by the Information Governance Review Panel (IGRP). 'Findings will be disseminated through peer reviewed publications, consultations with

stakeholders and national/ international conference presentations. The improved understanding of the prior health, nature of previous contacts with services, and wider social circumstances of those who complete suicide will assist in prevention policy, service organisation and delivery. SID-Cymru is funded through the National Institute for Social Care and Health Research, Welsh Government (RFS-12-14).

Article Summary

Article Focus:

- Case definitions of probable and possible suicides
- Identifying risk factors and patterns of service use in those who complete suicide
- Designing a study

Key messages:

- SID-Cymru will facilitate a series of case- control studies of those who complete suicide
- Recency of primary care, in- and out-patient hospital and emergency department contacts including attendance for self-harm, primary care diagnosis of depression, levels of treatment with anti-depressants and trends in such treatment over time, rural and urban geography, educational attainment, contacts for the elderly and levels of physical illness will be investigated
- Results will be adjusted for a number of confounders

Strengths and Limitations of this study:

- Based on the general population

- Includes all deaths through suicide, not just those known to mental health services
- Based on routinely collected health and social care data linkage
- Linkage across multiple domains. Most suicide data linkage studies are limited to only one or two domains of health care provision.
- Prescriptions not actual use

Introduction

Every year approximately 800,000 people die by suicide worldwide, one to two in every 100 deaths. Prevention of suicide is a global public health challenge. Collaborative working across government departments, with a public health approach extending beyond mental health service care is essential[1]. Global patterns and national trends in the incidence of suicide and its key risk factors change over time. The therapeutic and preventive challenges of understanding and responding to these changes are considerable.

Vast amounts of personal data are routinely collected on a daily basis by health and social care systems around the world to support clinical management and patient care. Linking these data records, for the same individuals across different services and over time, offers a powerful, population-wide resource. Such integrated datasets have been used to study a range of health issues to identify risk and protective factors and to examine outcomes. The secondary use of these data has enormous potential in suicide research. Improved consideration of the prior health, wider social circumstances and points of access to services of all individuals who complete suicide can be achieved.[2-4]

Studies from the Nordic countries, have demonstrated the usefulness of data-linkage across register based studies in suicide research.[5-10] Others[11-13] have demonstrated that collating and linking sets of routinely collected whole population based data, such as GP records, out-patient data, in-patient activity, with mortality data enable more detailed analysis of risk factors for those people completing suicide. Most previous research linking suicide mortality data with routinely collected electronic health records involves only one or two domains of health care provision such as psychiatric in-patient care.

In the United Kingdom various systems exist to examine suicide deaths. The National Confidential Inquiry into Suicide and Homicide (NCISH) by people with Mental Illness focuses on suicide cases that were in contact with mental health services (approximately 25% of total*)* in the year prior to their deaths.[3] However this provides limited information on issues of suicide in the general population and may hamper the effectiveness of wider preventive efforts.[1] In a recent report NCISH[14] examined aspects of primary healthcare prior to all suicides in England between 2002-2011; however no linkage was made with data from other service providers such as emergency departments.

Scotland has recently established enhanced data collection in relation to suicide, however, further development is needed in order for 'ScotSID' to be able to examine healthcare pathways and contact with more than one health service;[15] England does not currently have a dedicated repository for suicide data though studies have made use of the Clinical Practice Research Datalink (CPRD; www.cprd.com), which represents approximately 8.5% of the UK population from 600 General Practices (GP's) in England.[16] CPRD can be linked with data from the National Health Service (NHS) Hospital Episode Statistics (HES; http://www.hscic.gov.uk) and mortality data from the Office of National Statistics (ONS; http://www.ons.gov.uk) but has limited emergency department data.

SID-Cymru will access and link information on prior health, nature of previous contacts with services, and wider social circumstances of all those who complete suicide (known and unknown to mental health services) within the population of Wales via anonymised routinely collected electronic data held in healthcare and social datasets from the Secure Anonymised Information Linkage (SAIL) Databank (www.saildatabank.com).[17] [18] This brings together, links and anonymises the widest possible range of person-based data currently available in

the UK. The SAIL Databank was originally set up by the Health Information Research Unit (HIRU) at the College of Medicine at Swansea University.

SID-Cymru is part of the research programme related to the Health e-Research Collaboration UK (HeRC UK), led by the Medical Research Council (MRC) and based in the Centre for the Improvement of Population Health through e-Records Research (CIPHER). CIPHER is a UK Clinical Research Collaboration (UKCRC) Public Health Research Centre of Excellence set within the College of Medicine at Swansea University. SID-Cymru will provide timely robust data to inform the future strategic direction of and the first step in designing and evaluating effective interventions to prevent suicide. It will use the International Classification of Disease, Tenth Revision,[19] (ICD-10) definitions and instructions for classifying causes of death which will allow for comparisons with other countries and thus support research of relevance globally. An emphasis will be given to issues where there are opportunities for intervention or where electronic data linkage confers an advantage to its investigation.

Aim

The main aim is to establish a Suicide Information Database- Cymru as a population based resource for studying factors and service contacts associated with all suicide deaths through routinely collected data linkage case control studies.

Objectives:

Phase 1:

- To identify, via the SAIL Databank, all those with ICD-10 codes for probable and possible suicide in Wales 2003-2011, and matched controls.

- To explore and address the methodological issues relating to the development of this database of completed suicides and the linkage of data across different datasets and settings.

Phase 2:

- To investigate risk factors and trends for suicide including: primary care diagnosis of depression; levels of treatment with anti-depressants and trends in such treatment over time; rural and urban geography; educational attainment; levels of physical illness.
- To investigate settings and pathways of care where people are in contact with services in the year leading up to their suicide across the whole population and in specific groups such as the elderly.

Methods and Analyses

Design

SID-Cymru will facilitate a series of electronic population based, routinely collected data linkage case-control studies on completed suicide in Wales between 2003 and 2011. Wales has a population of 3.1 million[20] and is part of the United Kingdom. There are approximately 32,000 deaths of Welsh nationals registered each year of which around 300 (approximately 1%) are registered as suicides.[21]

Data Source

SID-Cymru will be assembled within the SAIL Databank,[17] a growing resource that already holds over a billion anonymised records from 13 databases which can be anonymously linked at the individual record level.[18] The SAIL Databank has been previously used for linkage of routine data.[22-26]

Within the SAIL Databank a split-file approach to anonymisation is used to overcome issues of confidentiality and disclosure in health-related data warehousing.[18] Demographic data are sent to a partner organisation, NHS Wales Information Service (NWIS), where all identifiable information is removed; clinical data are sent directly to HIRU, where, for each dataset within the SAIL Databank, an individual is assigned an encrypted Anonymised Linking Field (ALF). The ALF is used to link anonymised individuals across datasets, thus supporting the opportunity to conduct longitudinal analyses of an individual's journey through multiple health, education and social datasets.[17] Additionally Residential Anonymised Linking Fields (RALFs) have been created for all residences in Wales and enables linkage of anonymised household and environment data with the health records of individual residents without the identity of the residences or residents being known to researchers.[26]

The primary study base will be the Welsh Demographic Service[27] (WDS) hosted within the SAIL Databank. The WDS is a core dataset available within the SAIL Databank and part of a set of services to manage administrative information (demographic data) for NHS patients in Wales. The WDS was introduced early in 2009 replacing a similar service known as the NHS Wales Administrative Register (NHS AR). The WDS data is collected from GP practices via the Exeter System and more than five million individuals are currently present in the WDS dataset within the SAIL Databank. The WDS is a register of all individuals that have at some point in time been registered with a Welsh General Practitioner (GP) or required some form of NHS health care provision in Wales. The electronic collation of WDS / NHS AR data originated in 1960 and is updated and maintained by NWIS,[27] ensuring that address changes (within and out of Wales) and death notices are included in the register. The original (non-anonymised) version of the NHS AR has been used in the HIRU Matching Algorithm for Consistent Results in Anonymised Linkage (MACRAL), making the WDS / NHS AR the

master list for all Welsh residents and using probabilistic matching to find the associated NHS numbers that are then encrypted into ALF's.

Deaths in Wales should be registered within five days of the date of death. However legislation in Wales means that when a coroner's inquest takes place, the death cannot be registered until the inquest is complete. Since the Office for National Statistics (ONS), the national agency where all deaths are collated, has no conclusive information about the death until it is registered, there is a delay between the date the death occurred and when the death is added to the annual ONS mortality dataset. The ONS found that less than 41% of deaths going to inquest were registered within 3 months, though 96% were registered within 1 year.[28] Thus precise information on the annual incidence of suicide may be delayed by up to two years. Information collected at death registration is recorded on the Registration Online system by registrars. Most of the information is normally supplied by the informant (usually a close relative of the deceased) while the cause of death (COD) is usually obtained from the Medical Certificate of Cause of Death (MCCD) completed by a medical practitioner when the death is certified using ICD-10 coding or the coroner if there is an inquest.[19] Notably, a death is not officially registered within the Annual District Deaths Extract (ADDE) until the COD has been finalised, and thus the year of death and the year of registration may not concur. The *primary dataset* used to construct SID-Cymru is the ADDE from the ONS. The ADDE is inclusive of Welsh residents who died outside of Wales, and holds information about COD derived from death certificates on all deaths in Wales.

Definition of Suicide for Cases

The true number of suicides is difficult to determine because a coroner's conclusion of suicide must be 'beyond a reasonable doubt' i.e. that the death was intentionally self inflicted and in some areas coroners have increasingly (since 2001) reported narrative conclusions

rather than reporting it as suicide.[29 30] Previously when insufficient information was recorded by the coroner ONS coders used to record the death as an accident which inevitably led to some suicides being classified as accidents or misadventure. The ONS has recently issued guidance on this issue following a coding practice review.[31] Current ONS practice includes deaths where intention is 'undetermined whether accidentally or purposefully inflicted'; thus deaths where there may be no intention to take life, such as in relation to injury or poisoning, are included in suicide figures by ONS. Currently there is no access to coroners' narrative verdicts within the SAIL Databank as a possible method for review of case inclusion.

There is evidence to suggest that a high proportion of deaths from poisoning and hanging that receive accidental verdicts, are found, when subjected to clinical review, to be suicides.[32] Such possible deaths through suicide will be included in SID-Cymru as an opportunity for further separate and combined analysis; thus the additional ICD-10 codes relating to 'accidental poisoning with prescribed drugs' (X40-X41, X43-X49) and 'accidental hanging' (W75-W76) may be used along with 'sequelae of external causes of morbidity and mortality' (Y87, Y87.2, Y89, Y89.9).[18] Thus with SID-Cymru we aim to establish a Suicide Information Database that can be a resource for studying factors associated with *all* probable, as well as, possible deaths through suicide; by including accidental hanging/strangulation and accidental poisoning, excluding, narcotics and psychodysleptics (i.e., possible suicides), along with the ONS traditional method of defining suicides, which includes suicides and deaths of undetermined intent (probable suicides).[2]

Identification of Cases and Controls

SID-Cymru cases encompass suicides ('intentional self-harm' [ICD-10: X60-84]) and probable deaths through suicide ('undetermined intent' [ICD-10: Y10-34, excl.Y33.9]) recorded from the MCCD and presented within ADDE as the underlying COD. Those coded

Y33.9/U50.9 (pending verdicts) are excluded since a large proportion of these are subsequently found to be homicides. Probable suicide defined above can be supplemented with possible deaths through suicide if required. Cases of probable suicide will be identified and extracted by use of the ICD-10 codes[19] defined above and depicted in Table 1.

ONS figures include those over 15 years of age only, due to the possibility that deaths in younger children coded as undetermined events may be caused by unverifiable accidents, neglect or abuse.[33] SID-Cymru will allow for the analysis of suicides and probable suicides for the 10 to 14 year age band. Official ONS mortality statistics are produced based on the number of deaths registered in a particular calendar year, rather than the number of deaths that occurred in that year. This means their figures include some deaths that occurred in years prior to the reference year (approximately 4%). As SID-Cymru will link and review data in the period leading up to death it is important to ensure the match/end date reflects the correct time period (specific to each individual case), i.e. date of death not date of registration, to afford an accurate perspective on utilisation of resources and help seeking behaviours. Consequently the actual date of death (ADOD) will be used in the matching criteria to establish a data review 'end date' for controls, rather than the registered date of death (RDOD) referred to in ONS reports. Mortality data within the SAIL Databank is only available from 2003, thus the earliest case inclusion relates to ADOD's from the 1st January 2003. To minimise the underestimation of cases identified (i.e., due to delays in COD confirmation and registration of DOD) the case inclusion for SID-Cymru includes cases aged 10 years and over at ADOD, where the ADOD took place between 01/01/2003 and 31/12/2011.

Table 1. ICD-10 codes used to identify and extract cases with death through probable suicide for SID-Cymru.

Short name	**Description**	**ICD10 codes**
Poisoning	Poisoning by and exposure to one or more types of poisons	X60-X69, Y10-Y19
Hanging & suffocation	Hanging, strangulation and suffocation	X70, Y20
Drowning	Drowning and submersion	X71, Y21
Guns	Firearm discharge	X72-X74, Y22-Y24
Sharp object	Sharp object	X78, Y28
Jumping from a high place	Jumping from a high place	X80, Y30
Jumping or lying before moving object	Jumping or lying before moving object	X81, Y31
Explosives & Fire	Explosive material, smoke, fire, flames and hot objects	X75-X77, Y25-Y27
Other or unspecified events	Intentional self-harm by other/un- specified means; Other/un- specified events undetermined intent	X83-X84, Y33-Y34 (excl.Y33.9)
Crash by motor vehicle	Crashing of motor vehicle intentional and undetermined intent	X82, Y32
Blunt Object	Blunt Object	X79, Y29

Table 2 presents the number of registered deaths through suicide per year and actual deaths per year between 2003 and 2011 as reported by the ONS and as identified for SID-Cymru within the SAIL Databank.

Table 2. Probable Suicide[1] Deaths for Wales 2003 – 2011 identified within the SAIL Databank for SID-Cymru.

Count by Year[2]	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**TOTAL**
Probable Suicides ADOD[3] within SID[4]	361	303	287	278	279	282	272	292	304	**2664**
Probable Suicides RDOD[5] within SID	274	330	270	300	288	266	257	287	340	**2612**
ONS Probable Suicides[6] RDOD	332	330	269	300	289	266	258	288	341	**2673**

[1] Probable suicides = ICD-10 codes for Intentional [X60-84], Undetermined [Y10-34 excl.Y33.9] self-harm; [2] Values are number of individuals unless otherwise indicated [3] ADOD = Year death occurred; [4] Cases identified for SID-Cymru includes those aged 10 years and over; [5] RDOD = Year death was registered; [6] Cases reported by the ONS include those aged 15 years and over.

Matched Controls will be identified within the WDS as live individuals matched on age (to the nearest year) and gender who were registered within the WDS for at least one year prior to matched case's ADOD. The controls will be required to be alive at the time of the matched case's ADOD (i.e., the match/end date). The use of live controls limits the introduction of bias relating to deaths, particularly in the younger age groups, of those undertaking risky behaviours resulting in premature death (e.g. substance use, accidents) that may be associated with unrecognised suicidal behaviours or known risk factors. To add to the power of the study we aim to identify at least five controls to every one case.[34]

Routine Data Sources

For SID-Cymru the data collected on identified cases and controls, via ADDE and the WDS respectively, will be linked to other routinely collected datasets, retrospectively allowing a review of each individual's pathway through the various services. Linkage with, for example, GP system data provides varying information about patients going back several years, including previous diagnosis, presenting symptoms and previous medications prescribed. This dataset can be used to review contacts with the GP and, consequently, infer the development / diagnosis of any new medical conditions including depression and self harm prior to suicide. Linkage with in-patient data will allow a review of hospital contacts and Emergency Department datasets will give information on crisis contacts. This will provide comprehensive insights into help seeking behaviours and management across settings. Data sets currently accessible via the SAIL Databank which will be linked to SID-Cymru are presented in Table 3. There are several other datasets currently under negotiation for inclusion within the SAIL Databank to which, in the future, SID-Cymru could link. These include Department of Work and Pensions (DWP) employment and incapacity status, Looked After Children, fostering, Substance Misuse Services, Sexually Transmitted Infections and Police Data.

Table 3. Datasets available within the SAIL Databank for linkage with SID-Cymru cases and controls.

Dataset	Data Description	Data Provider	Data Type	Geographical Coverage	Available from
Annual District Death Extract (ADDE)	Register of all deaths relating to Welsh residents, including those that died out of Wales.	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW) on behalf of ONS	Registry data	Wales	2003
Welsh Demographic Service (WDS)	Administrative information about individuals in Wales that use NHS services; such as address and practice registration history. It replaced the NHS Wales Administrative Register (NHSAR) in 2009.	NHS Wales' Informatics Service (NWIS)/ HSW	Administrative (mandatory) data return	Wales	1990
Area Based Deprivation Scores	Lower Super Output Area - Townsend Deprivation Index and Welsh Index of Multiple Deprivation Scores.	NHS Wales' Informatics Service (NWIS)/ HSW	Administrative data	Wales	2001 / 2005
Education Attainment (DCEL)	Educational attainment at key stages (pre 16 year old). Data collated from school census, the national pupil survey and educational attainment databases by the Welsh Government.	Department for Children, Education, Lifelong Learning and Skills (DCELLS)	Administrative / survey data.	Wales	2005
Primary Care GP dataset	Each GP practice uses a clinical information system to maintain an electronic health record for each of their patients; capturing the signs, symptoms, test results, diagnoses, prescribed treatment, referrals for specialist treatment and social aspects relating to the patients home environment.	Welsh General Practices that have signed up to the SAIL Databank	Clinical system data	Wales	2000
Patient Episode Database for Wales (PEDW)	NHS Wales - hospital admissions (Inpatients and day cases) dataset comprising of attendance and clinical information for all hospital admissions: includes diagnoses and operations performed.	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW)	Administrative (mandatory) data return	Wales	1997
Outpatient Dataset (OPD)	Attendance information for all NHS Wales hospital outpatient appointments.	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW)	Administrative (mandatory) data return	Wales	2004
Emergency Department Data Set (EDDS)	Administrative and clinical information for all NHS Wales Accident and Emergency department attendances. Includes the All Wales Injury Surveillance Systems (AWISS) dataset	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW)	Administrative (mandatory) data return	Wales	2009

Measurements

Data Variables/Characteristics

Data to be included in SID-Cymru will be extracted from the SAIL Databank and will include basic demographics; educational data; ADOD and RDOD; numbers and percentages for deprivation; proportions known to different healthcare settings in the period prior to death; inclusive medical history, i.e., primary care contact and diagnosis (by Read Codes), information about hospital / psychiatric admission and diagnosis (by ICD-10 Codes), and nature of service contact e.g. for self-harm, substance misuse. Therefore, the routinely collected data held by the NHS and other public bodies supplying the existing SAIL Databank will maximise the narrative of a death through suicide while being less resource intensive than psychological autopsies. The initial variables to be extracted and linked across data sets are described in Table 4.

Planned Analyses

The primary objective is to establish SID-Cymru as a resource for future analysis. It is important to maximise the utility of the resource, and some general principles of analysis have been determined.

Descriptive Epidemiology

Phase 1:

- Identification and description of Cases: Number of deaths registered with ONS and available within the SAIL Databank, with relevant ICD-10 codes defined as suicides, between 2003-2011.

Table 4. Characteristics to be identified and collected through data linkage for SID-Cymru Cases and Controls.

Data	**Characteristics**
Demographics	Sex
	Date of Birth
	Marital status*
	Employment*
	Education*
Mortality	Date of Death,
	Age at DOD (in years)
	Registered Date of Death
	Cause of Death (ICD-10 code)
Deprivation	LSOA
	WIMD
	Townsend
	Rural-Urban Indicators
Medical History	GP register*
	GP diagnosis (Read code)*
	GP events*
	Hospital attendance
	Hospital diagnosis
	Hospital events
	Outpatient attendance
	Outpatient diagnosis*
	Outpatient events
	ED attendance*
	ED diagnosis*
	ED events*
	Psychiatric Admission*

* Where data are available; there may be limitations to the number of cases with this information.

- Identification and description of Matched Controls.
- Basic demographics.
- Table of delay (days) in registering suicides and undetermined deaths.

Phase 2:

- Proportions known to different healthcare settings: number and percentages, with main diagnosis, that had a general hospital admission; emergency department contact for self-harm and other indications; psychiatric admission and primary care contact in the year prior to probable suicide.
- Numbers and percentages for deprivation, employment status, educational achievement and medical history (e.g., chronic pain, terminal illness, medication, previous self harm, and substance misuse) will also be sought.
- Number of cases with missing data across data sets for variables of interest will be noted.

Area based measures of socio-economic deprivation

Deprivation will be measured at Lower Super Output Area (LSOA) level using the Welsh Index of Multiple Deprivation[35] (WIMD) and Townsend Index Score.[36] All suicides and matched controls will be assigned to a lower layer super output area (LSOA). There are 1909 LSOAs in Wales with an average population of 1500 people (range: 1000-3000).[37] Linkage to WIMD and Townsend Index information is available in the SAIL Databank. These will be ranked for deprivation, divided into quintiles and standardised rates calculated.

Odds ratios for the described exposures in the case control study

A case-control study utilising SID-Cymru will be population based and so the relative risk of suicide will be estimated by conditional logistic regression model with SPSS (version 20). Crude odds ratios will be adjusted for general practice and/or LSOA by matching cases and controls. Unadjusted estimates, confounder-adjusted estimates and their precision (e.g., 95% confidence interval) will be produced. Interactions between variables will be assessed with the log likelihood ratio test based on results from the adjusted analysis. The population attributable risk will be calculated[38] on the basis of adjusted relative risks from the full analysis and the distribution of exposures in the cases.

We will also report information on the completeness of linkage with each dataset.

Ethics and Dissemination

Ethics

A large amount of preliminary work on anonymisation methodologies was undertaken to create the SAIL Databank system,[17- 18] and the SAIL Databank has the required ethical permissions and processes in place to analyse anonymised data. It operates within a robust series of guidelines in line with the Caldicott principles and the National Information Governance Board for Health and Social Care.[18] Ethical approval has been granted for SID-Cymru from HIRU's Information Governance Review Panel (IGRP), an independent body consisting of a range of government, regulatory and professional agencies. In compliance with the IGRP rulings and the Data Protection Act 1998[39] individual level data and personal identifier linkage codes will not be removed from the SAIL Databank and all analyses will be carried out within the SAIL Databank gateway at Swansea University, a secure access point to data within the SAIL Databank. The key points of the MRC/Wellcome Trust data sharing policy will be followed.[40 41]

Dissemination

This paper describes the protocol for the development of a Suicide Information Database (SID-Cymru), and the research opportunities available from an electronic case-control study of suicides within a whole population. SID-Cymru will have the ability to link suicide cases anonymously to primary and secondary health information along with other social care data, allowing us to review each case's journey through these datasets. The establishment of SID-Cymru and exploration of the linkage methodologies will improve our understanding of those who complete suicide (particularly those not known to mental health services) and will be used to inform service planning and policy decision making and implementation. It will help identify key opportunities and settings for prevention of this tragic event. By so doing, SID-Cymru will join other international databases of suicide research and provide a platform for further investigation and data linkages.

In order for SID-Cymru to become a functional resource it is important to be aware of the limits of health data available; though widely used in research, and offering a broad range of information about treatment and associated conditions, there are issues relating to determining the quality of patient records, the completeness of data available and any conclusions that may be drawn from them, perhaps particularly concerning primary care records.[42] That is, working with routinely collected data presented in its "raw" format requires a certain level of database analysis skills where duplicates, missing and erroneous entries are common occurrences. Whilst some such administrative / system-based recording issues are easy to identify and account for in individual datasets, it is not always apparent what is correct and what is erroneous at the combined level. Indeed, this problem is confounded when linkage of data reveals conflicting information causing routine data to appear inaccessible and attempts at linkage discouraged. Thus, a secondary aim for SID-Cymru is to share the skills developed as part of establishing a suicide database, which can

aid colleagues who may lack such analytical expertise and foster greater multi-disciplinary collaborations and advance suicide research.

The UK has a strong presence in form of a wide range of publications and expertise relating to suicide research. Successful and dedicated Suicide Research Centres exist in Bristol, Manchester and Oxford,[43-45] and Scotland recently commenced work on a Scottish Suicide Information Database ('ScotSID').[46] These centres of excellence report broadly about suicide, though for logistical reasons, often a regional focus is retained, relating to project funding, analytical and interventional work; for that reason the establishment of a Suicide Information Database in Wales will build on and enhance existing UK suicide resources and infrastructure since the level of linkage available within the SAIL Databank is unique in this field. Unique opportunities within the SAIL Databank include linkage across primary, secondary and emergency department data and with education data. Findings will be disseminated through publications in peer reviewed journals and presentations at local, national and international conferences. Communication and consultation with key stakeholders from health and social care (for example primary care, mental health, Royal Colleges), government and other policy makers, as well as, the third sector will occur. Dissemination will be facilitated by the wider roles and responsibilities in suicide prevention, nationally and internationally, of members of the research team.

Implications and Significance

Valuable opportunities exist for a wide range of epidemiological and clinical studies on suicide in Wales and SID-Cymru has the potential to become an important resource in facilitating such research which will be of relevance internationally. In addition to the records that have already been included for linkage with SID-Cymru it is expected that over time relevant information from other data sources (e.g. the DWP) will be linked to SID-Cymru to provide a wider range of information on issues such as individuals' circumstances, the nature of their deaths, and their contact with extended services. Additionally, linkage can be made with non-routinely collected datasets such as that held by NCISH. Specific hypotheses that will be explored include: recency of primary care, hospital and emergency department contacts including attendance for self-harm, primary care diagnosis of depression, levels of treatment with anti-depressants and trends in such treatment over time, rural and urban geography, contacts for the elderly and levels of physical illness.

Findings of current and projected public health importance will be assessed and presented to support policy makers, commissioners and providers of health and social care in Wales. Non-identifiable information from this project will be made available to researchers in Wales, the UK and international collaborators. The initial project focus will be on identification of cases and controls, data linkage opportunities and methodological issues relating to the establishment of a Suicide Information Database and routine data linkage (Phase 1), before commencing data extraction and analysis (Phase 2). The proposed data collation and linkage of primary, secondary and emergency department health information together with educational data are currently unique in the United Kingdom. Thus it is possible to develop a central repository for information relating to suicide in whole population that will be of relevance internationally. This paper describes the design and development of a Suicide Information Database in its infancy.

Authors' contributions

AJ, the principal investigator, was responsible for the conceptualization of SID-Cymru. AJ, KL, MD, DG, JS for the design of SID-Cymru; AJ and LK for its on-going operationalisation. AJ and LK drafted the manuscript. All authors read and approved the final manuscript.

Funding Statement

This study was funded by a grant from the National Institute for Social Care and Health Research, Welsh Government, grant number RFS-12-14. DG is a National Institute for Health Research (England) Senior Investigator.

Competing interests

The authors declare that they have no competing interests.

Ethics approval

Ethical approval has been granted by the Health Information Research Unit Information Governance Review Panel at the College of Medicine at Swansea University.

Data sharing statement

It will be possible to access the data after the publication of the results. Researchers interested in collaborations or further information are invited to contact AJ at a.john@swansea.ac.uk.

References

1. World Health Organisation. 66th World Health Assembly. Draft comprehensive mental health action plan 2013-2020. A66/10 Rev.1, WHO 16 May 2013. http://apps.who.int/gb/e/e_wha66.html Accessed September 2014.
2. Linsley KR, Schapira K, Kelly TP. Open verdict v. suicide—importance to research. *Br J Psychiatry* 2001, **178**:465-468.
3. Appleby L, Kapur N, Shaw J, *et al*. Annual Report: England, Northern Ireland, Scotland and Wales. Manchester: University of Manchester 2013.

http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/AnnualReport2013_UK.pdf Accessed July 2014.

4. Andersen UA, Andersen M, Rosholm JU, *et al*. Contacts to the health care system prior to suicide: a comprehensive analysis using registers for general and psychiatric hospital admissions, contacts to general practitioners and practising specialists and drug prescriptions. *Acta Psychiatr Scand* 2000, **102**:126-34.
5. Qin P, Agerbo E, Mortensen PB. Suicide risk in relation to socioeconomic, demographic, psychiatric, and familial factors: a national register–based study of all suicides in Denmark, 1981–1997. *Am J Psychiatry* 2003, **160**:765-772.
6. Gunnell D, Löfving S, Gustafsson JE, *et al*. School performance and risk of suicide in early adulthood: follow-up of two national cohorts of Swedish schoolchildren. *Journal of affective disorders*, 2011, **131**:104-112.
7. Qin P, Hawton K, Mortensen PB, *et al*. Combined effects of physical illness and comorbid psychiatric disorder on risk of suicide in a national population study. *Br J Psychiatry* 2014, **204**:430-435.
8. Christiansen E, Larsen KJ, Agerbo E, *et al*. Incidence and risk factors for suicide attempts in a general population of young people: a Danish register-based study. *Aust N Zealand J Psychiatry* 2012, **47**: 259-270.
9. Björkenstam C, Johansson LA, Nordström P, *et al*. Suicide or undetermined intent? A register-based study of signs of misclassification. *Population health metrics*, 2014, **12**:11.
10. Hjorthøj CR, Madsen T, Agerbo E, *et al*. Risk of suicide according to level of psychiatric treatment: a nationwide nested case–control study. *Social psychiatry and psychiatric epidemiology*, 2014; 1-9.

11. Lawrence DM, Holman CD, Jablensky AV, *et al*. Suicide rates in psychiatric in-patients: an application of record linkage to mental health research**.** *Aust NZ J Public Health* 1999, **23**:468–470.

12. Lawrence D, D'Arcy C, Holman J, *et al.* Increasing rates of suicide in Western Australian psychiatric patients: a record linkage study. *Acta Psychiatr Scand* 2001, **104**:443-451.

13. Morrison KB, Laing L. *Adults' use of health services in the year before death by suicide in Alberta.* Statistics Canada; 2011.

14. Appleby L, Kapur N, Shaw J, *et al*. Suicide in primary care in England: 2002-2011. National Confidential Inquiry into Suicide and Homicide by People with Mental Illness (NCISH). Manchester: University of Manchester 2014. http://www.bbmh.manchester.ac.uk/cmhr/research/centreforsuicideprevention/nci/reports/SuicideinPrimaryCare2014.pdf Accessed August 2014.

15. ISD Scotland: The Scottish Suicide Information Database Report 2014. http://www.isdscotland.org/Health-Topics/Public-Health/Publications/index.asp?ID=1232 Accessed August 2014.

16. Webb RT, Kontopantelis E, Doran T, *et al*. Suicide risk in primary care patients with major physical diseases: a case-control study. *Arch Gen Psychiatry* 2012, **69**:256-264.

17. Lyons RA, Jones KH, John G, *et al*. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Making* 2009, **9**:3.

18. Ford DV, Jones KH, Verplancke JP, *et al*. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009, **9**:1.

19. World Health Organization: *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. WHO: Geneva; 1993.

20. Office of National Statistics. 2011 Census Results for Wales. ONS 16th July 2012. http://wales.gov.uk/docs/statistics/2012/120716-census-2011-first-results-wales-en.pdf Accessed September 2014.

21. Office of National Statistics. Statistical bulletin: Suicide rates in the United Kingdom, 2006 to 2010. ONS 26th January 2012. http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2010/stb-statistical-bulletin.html Accessed August 2014.

22. McGregor JI, Brooks CJ, Chalasani P, *et al*. The Health Informatics Trial Enhancement Project (HITE): Using routinely collected primary care data to identify potential participants for a depression trial. *Trials* 2010, **11**:39.

23. Brooks CJ, Stephens JW, Price DE, *et al*. Use of a patient linked data warehouse to facilitate diabetes trial recruitment from primary care. *Primary Care Diabetes* 2009, **3**:245-248

24. Hill RA, Brophy S, Brunt H, *et al*. Protocol of the baseline assessment for the Environments for Healthy Living (EHL) Wales cohort study. *BMC Public Health* 2010, **10**:150

25. Fone DL, Dunstan F, White J, *et al*. Cohort Profile: The Caerphilly Health and Social Needs Electronic Cohort Study (E-CATALyST). *International Journal of Epidemiology*, 2013, **42**:1620-1628.

26. Rodgers SE, Heaven M, Lacey A, et al. Cohort profile: the housing regeneration and health study. *Int J Epidemiol* 2012, dys200.

27. SAIL Databank: Core Data Sets. Swansea University: Swansea http://www.saildatabank.com/data-dictionary/sail-datasets Accessed September 2014.

28. Brock A, Baker A, Griffiths C, *et al*. Suicide trends and geographical variations in the United Kingdom, 1991–2004. *Health Statistics Quarterly* 2006, **31**:6-22.

29. Carroll R, Hawton K, Kapur N, *et al*. Impact of the growing use of narrative verdicts by coroners on geographic variations in suicide: analysis of coroners' inquest data. *Journal of Public Health* 2011, **10**:1-7.

30. Gunnell D, Hawton, K, Kapur N. Coroners' verdicts and suicide statistics in England and Wales. *BMJ* 2011, **343**:d6030.

31. Hill C, Cook, L. Narrative verdicts and their impact on mortality statistics in England and Wales. *Health Statistics Quarterly* 2011, **49**:81-100.

32. Gunnell D, Bennewith O, Simkin S, *et al*. Time trends in coroners' use of different verdicts for possible suicides and their impact on officially reported incidence of suicide in England: 1990-2005. *Psychol Med* 2013, **43**:1415-1422.

33. Office of National Statistics. Statistical bulletin: Suicide rates in the United Kingdom, 2011. ONS 22 January 2013. http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2011/stb-suicide-bulletin.html Accessed August 2014.

34. Grimes DA, Schultz KF. Compared to what? Finding controls in a case control study. *Lancet* 2005, **365**: 1429–33.

35. Office for National Statistics. Welsh index of Multiple Deprivation 2011. Summary report, Welsh Government; 2011.

36. Townsend P, Phillimore P, Beattie A. *Health and Deprivation: Inequality and the North*. Routledge: London; 1988.

37. Office for National Statistics. Super Output Areas 2011. http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html Accessed August 2014.

38. Bruzzi P, Green SB, Byar DP, *et al*. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol* 1985, **122**:904-914.

39. Great Britain: Data Protection Act 1998. London: HMSO. http://www.legislation.gov.uk/ukpga/1998/29/contents Accessed August 2014.

40. MRC: Policy and Guidance on Sharing of Research Data from Population and Patient Studies. MRC 23rd November 2011. http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC008302 Accessed August 2014.

41. Wellcome Trust: Data sharing policy. Wellcome Trust August 2010. http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm Accessed August 2014.

42. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ* 2003, **326**:1070.

43. University of Bristol, Suicide Research Centre. http://www.bristol.ac.uk/psychiatry/research/suicideresearch.html Accessed August 2014.

44. The University of Manchester Centre for Suicide Prevention. http://www.bbmh.manchester.ac.uk/cmhr/research/centreforsuicideprevention/ Accessed August 2014.

45. University of Oxford, Centre for Suicide Research. http://cebmh.warne.ox.ac.uk/csr/ Accessed August 2014.

46. ISD Scotland. Scottish Suicide Information Database (ScotSID). ISD Scotland 2011. http://data.gov.uk/dataset/scottish_suicide_information_database Accessed August 2014.

Suicide Information Database- Cymru: A protocol for a population-~~-~~based, routinely collected data linkage ~~case control~~ study to explore risks and patterns of healthcare contact prior to suicide to identify opportunities for intervention.

John A[1*], Dennis M[1,] Kosnes L[1], Gunnell D[2], Scourfield J[3], Ford DV[1], Lloyd K[1]

[1]College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park, Swansea, SA2 8PP, United Kingdom.

[2]School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, United Kingdom.

[3]School of Social Sciences, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff, CF10 3WT, United Kingdom.

Corresponding Author: Ann John[1] a.john@swansea.ac.uk; Phone: +44(0)1792 602568; College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park, Swansea, SA2 8PP, United Kingdom.

Michael Dennis[1] m.s.dennis@swansea.ac.uk, Liv Kosnes[1] l.kosnes@swansea.ac.uk, David Gunnell[2] d.j.gunnell@bristol.ac.uk , Jonathan Scourfield[3] Scourfield@cardiff.ac.uk, David V. Ford[1] d.v.ford@swansea.ac.uk, Keith Lloyd[1] k.r.lloyd@swansea.ac.uk

Key words: Suicide, case-control, database, routine data linkage.

Word Count: 4054

Abstract

Introduction

Prevention of suicide is a global public health challenge extending beyond mental health services. Linking routinely collected health and social care system data records, for the same individual across different services and over time has enormous potential in suicide research. Most previous research linking suicide mortality data with routinely collected electronic health records involves only one or two domains of health care provision such as psychiatric in-patient care. This protocol paper describes the development of a population based, routinely collected data linkage study: the Suicide Information Database (SID-Cymru). SID-Cymru aims to contribute to the information available on people who complete suicide.

Methods and Analysis

SID-Cymru will facilitate a series of electronic case-control studies based in the Secure Anonymised Information Linkage (SAIL) Databank. We have identified 2664 cases of suicide in Wales between 2003 and 2011 from routinely collected mortality data using International Classification of Diseases, Tenth Revision, codes X60-84 (intentional self harm) and Y10-34 (undetermined intent). Each case will be matched by age and sex to at least five controls~~ who are alive at the time of death of the case~~. Records will be collated and linked from routinely collected health and social data in Wales for each individual. Conditional logistic regression will be applied to produce crude and confounder (including general practice, socioeconomic status) adjusted odds ratios.

Ethics and Dissemination

The SAIL Databank has the required ethical permissions in place to analyse anonymised data. Ethical approval has been granted by the Information Governance Review Panel (IGRP).

'Findings will be disseminated through peer reviewed publications, consultations with stakeholders and national/ international conference presentations. The improved understanding of the prior health, nature of previous contacts with services, and wider social circumstances of those who complete suicide will assist in prevention policy, service organisation and delivery. SID-Cymru is funded through the National Institute for Social Care and Health Research, Welsh Government (RFS-12-14).

Formatted: Font: Times New Roman, 12 pt

Article Summary

Article Focus:

- Case definitions of probable and possible suicides
- Identifying risk factors and patterns of service use in those who complete suicide
- Designing a study

Key messages:

- SID-Cymru will facilitate a series of case- control studies of those who complete suicide
- Recency of primary care, in- and out-patient hospital and emergency department contacts including attendance for self-harm, primary care diagnosis of depression, levels of treatment with anti-depressants and trends in such treatment over time, rural and urban geography, educational attainment, contacts for the elderly and levels of physical illness will be investigated
- Results will be adjusted for a number of confounders

Strengths and Limitations of this study:

- Based on the general population
- Includes all deaths through suicide, not just those known to mental health services
- Based on routinely collected health and social care data linkage
- Linkage across multiple domains. Most suicide data linkage studies are limited to only one or two domains of health care provision.
- Prescriptions not actual use

Introduction

Every year approximately 800,000 people die by suicide worldwide, one to two in every 100 deaths. Prevention of suicide is a global public health challenge. Collaborative working across government departments, with a public health approach extending beyond mental health service care is essential[1]. Global patterns and national trends in the incidence of suicide and its key risk factors change over time. The therapeutic and preventive challenges of understanding and responding to these changes are considerable.

Vast amounts of personal data are routinely collected on a daily basis by health and social care systems around the world to support clinical management and patient care. Linking these data records, for the same individuals across different services and over time, offers a powerful, population-wide resource. Such integrated datasets have been used to study a range of health issues to identify risk and protective factors and to examine outcomes. The secondary use of these data has enormous potential in suicide research. Improved consideration of the prior health, wider social circumstances and points of access to services of all individuals who complete suicide can be achieved.[2-4]

Studies from the Nordic countries, have demonstrated the usefulness of data-linkage across register based studies in suicide research.[5-10] Others[11-13] have demonstrated that collating and linking sets of routinely collected whole population based data, such as GP records, out-patient data, in-patient activity, with mortality data enable more detailed analysis of risk factors for those people completing suicide. Most previous research linking suicide mortality data with routinely collected electronic health records involves only one or two domains of health care provision such as psychiatric in-patient care.

In the United Kingdom various systems exist to examine suicide deaths. The National Confidential Inquiry into Suicide and Homicide (NCISH) by people with Mental Illness focuses on suicide cases that were in contact with mental health services (approximately 25% of total*)* in the year prior to their deaths.[3] However this provides limited information on issues of suicide in the general population and may hamper the effectiveness of wider preventive efforts.[1] In a recent report NCISH[14] examined aspects of primary healthcare prior to all suicides in England between 2002-2011; however no linkage was made with data from other service providers such as emergency departments.

Scotland has recently established enhanced data collection in relation to suicide, however, further development is needed in order for 'ScotSID' to be able to examine healthcare pathways and contact with more than one health service;[15] England does not currently have a dedicated repository for suicide data though studies have made use of the Clinical Practice Research Datalink (CPRD; www.cprd.com), which represents approximately 8.5% of the UK population from 600 General Practices (GP's) in England.[16] CPRD can be linked with data from the National Health Service (NHS) Hospital Episode Statistics (HES; http://www.hscic.gov.uk) and mortality data from the Office of National Statistics (ONS; http://www.ons.gov.uk) but has limited emergency department data.

SID-Cymru will access and link information on prior health, nature of previous contacts with services, and wider social circumstances of all those who complete suicide (known and unknown to mental health services) within the population of Wales via anonymised routinely collected electronic data held in healthcare and social datasets from the Secure Anonymised Information Linkage (SAIL) Databank (www.saildatabank.com).[17 18] This brings together, links and anonymises the widest possible range of person-based data currently available in

the UK. The SAIL Databank was originally set up by the Health Information Research Unit (HIRU) at the College of Medicine at Swansea University.

SID-Cymru is part of the research programme related to the Health e-Research Collaboration UK (HeRC UK), led by the Medical Research Council (MRC) and based in the Centre for the Improvement of Population Health through e-Records Research (CIPHER). CIPHER is a UK Clinical Research Collaboration (UKCRC) Public Health Research Centre of Excellence set within the College of Medicine at Swansea University. SID-Cymru will provide timely robust data to inform the future strategic direction of and the first step in designing and evaluating effective interventions to prevent suicide. It will use the International Classification of Disease, Tenth Revision,[19] (ICD-10) definitions and instructions for classifying causes of death which will allow for comparisons with other countries and thus support research of relevance globally. An emphasis will be given to issues where there are opportunities for intervention or where electronic data linkage confers an advantage to its investigation.

Aim

The main aim is to establish a Suicide Information Database- Cymru as a population based resource for studying factors and service contacts associated with all suicide deaths through routinely collected data linkage case control studies.

Objectives:

Phase 1:

- To identify, via the SAIL Databank, all those with ICD-10 codes for probable and possible suicide in Wales 2003-2011, and matched controls.

- To explore and address the methodological issues relating to the development of this database of completed suicides and the linkage of data across different datasets and settings.

Phase 2:

- To investigate risk factors and trends for suicide including: primary care diagnosis of depression; levels of treatment with anti-depressants and trends in such treatment over time; rural and urban geography; educational attainment; levels of physical illness.
- To investigate settings and pathways of care where people are in contact with services in the year leading up to their suicide across the whole population and in specific groups such as the elderly.

Methods and Analyses

Design

SID-Cymru will facilitate a series of electronic population based, routinely collected data linkage case-control studies on completed suicide in Wales between 2003 and 2011. Wales has a population of 3.1 million[20] and is part of the United Kingdom. There are approximately 32,000 deaths of Welsh nationals registered each year of which around 300 (approximately 1%) are registered as suicides.[21]

Data Source

SID-Cymru will be assembled within the SAIL Databank,[17] a growing resource that already holds over a billion anonymised records from 13 databases which can be anonymously linked at the individual record level.[18] The SAIL Databank has been previously used for linkage of routine data.[22-26]

Within the SAIL Databank a split-file approach to anonymisation is used to overcome issues of confidentiality and disclosure in health-related data warehousing.[18] Demographic data are sent to a partner organisation, NHS Wales Information Service (NWIS), where all identifiable information is removed; clinical data are sent directly to HIRU, where, for each dataset within the SAIL Databank, an individual is assigned an encrypted Anonymised Linking Field (ALF). The ALF is used to link anonymised individuals across datasets, thus supporting the opportunity to conduct longitudinal analyses of an individual's journey through multiple health, education and social datasets.[17] Additionally Residential Anonymised Linking Fields (RALFs) have been created for all residences in Wales and enables linkage of anonymised household and environment data with the health records of individual residents without the identity of the residences or residents being known to researchers.[26]

The primary study base will be the Welsh Demographic Service[27] (WDS) hosted within the SAIL Databank. The WDS is a core dataset available within the SAIL Databank and part of a set of services to manage administrative information (demographic data) for NHS patients in Wales. The WDS was introduced early in 2009 replacing a similar service known as the NHS Wales Administrative Register (NHS AR). The WDS data is collected from GP practices via the Exeter System and more than five million individuals are currently present in the WDS dataset within the SAIL Databank. The WDS is a register of all individuals that have at some point in time been registered with a Welsh General Practitioner (GP) or required some form of NHS health care provision in Wales. The electronic collation of WDS / NHS AR data originated in 1960 and is updated and maintained by NWIS,[27] ensuring that address changes (within and out of Wales) and death notices are included in the register. The original (non-anonymised) version of the NHS AR has been used in the HIRU Matching Algorithm for Consistent Results in Anonymised Linkage (MACRAL), making the WDS / NHS AR the

master list for all Welsh residents and using probabilistic matching to find the associated NHS numbers that are then encrypted into ALF's.

Deaths in Wales should be registered within five days of the date of death. However legislation in Wales means that when a coroner's inquest takes place, the death cannot be registered until the inquest is complete. Since the Office for National Statistics (ONS), the national agency where all deaths are collated, has no conclusive information about the death until it is registered, there is a delay between the date the death occurred and when the death is added to the annual ONS mortality dataset. The ONS found that less than 41% of deaths going to inquest were registered within 3 months, though 96% were registered within 1 year.[28] Thus precise information on the annual incidence of suicide may be delayed by up to two years. Information collected at death registration is recorded on the Registration Online system by registrars. Most of the information is normally supplied by the informant (usually a close relative of the deceased) while the cause of death (COD) is usually obtained from the Medical Certificate of Cause of Death (MCCD) completed by a medical practitioner when the death is certified using ICD-10 coding or the coroner if there is an inquest.[19] Notably, a death is not officially registered within the Annual District Deaths Extract (ADDE) until the COD has been finalised, and thus the year of death and the year of registration may not concur. The *primary dataset* used to construct SID-Cymru is the ADDE from the ONS. The ADDE is inclusive of Welsh residents who died outside of Wales, and holds information about COD derived from death certificates on all deaths in Wales.

Definition of Suicide for Cases

The true number of suicides is difficult to determine because a coroner's conclusion of suicide must be 'beyond a reasonable doubt' i.e. that the death was intentionally self inflicted and in some areas coroners have increasingly (since 2001) reported narrative conclusions

rather than reporting it as suicide.[29 30] Previously when insufficient information was recorded by the coroner ONS coders used to record the death as an accident which inevitably led to some suicides being classified as accidents or misadventure. The ONS has recently issued guidance on this issue following a coding practice review.[31] Current ONS practice includes deaths where intention is 'undetermined whether accidentally or purposefully inflicted'; thus deaths where there may be no intention to take life, such as in relation to injury or poisoning, are included in suicide figures by ONS. Currently there is no access to coroners' narrative verdicts within the SAIL Databank as a possible method for review of case inclusion.

There is evidence to suggest that a high proportion of deaths from poisoning and hanging that receive accidental verdicts, are found, when subjected to clinical review, to be suicides.[32] Such possible deaths through suicide will be included in SID-Cymru as an opportunity for further separate and combined analysis; thus the additional ICD-10 codes relating to 'accidental poisoning with prescribed drugs' (X40-X41, X43-X49) and 'accidental hanging' (W75-W76) may be used along with 'sequelae of external causes of morbidity and mortality' (Y87, Y87.2, Y89, Y89.9).[18] Thus with SID-Cymru we aim to establish a Suicide Information Database that can be a resource for studying factors associated with *all* probable, as well as, possible deaths through suicide; by including accidental hanging/strangulation and accidental poisoning, excluding, narcotics and psychodysleptics (i.e., possible suicides), along with the ONS traditional method of defining suicides, which includes suicides and deaths of undetermined intent (probable suicides).[2]

Identification of Cases and Controls

SID-Cymru cases encompass suicides ('intentional self-harm' [ICD-10: X60-84]) and probable deaths through suicide ('undetermined intent' [ICD-10: Y10-34, excl.Y33.9]) recorded from the MCCD and presented within ADDE as the underlying COD. Those coded

Y33.9/U50.9 (pending verdicts) are excluded since a large proportion of these are subsequently found to be homicides. Probable suicide defined above can be supplemented with possible deaths through suicide if required. Cases of probable suicide will be identified and extracted by use of the ICD-10 codes[19] defined above and depicted in Table 1.

ONS figures include those over 15 years of age only, due to the possibility that deaths in younger children coded as undetermined events may be caused by unverifiable accidents, neglect or abuse.[33] SID-Cymru will allow for the analysis of suicides and probable suicides for the 10 to 14 year age band. Official ONS mortality statistics are produced based on the number of deaths registered in a particular calendar year, rather than the number of deaths that occurred in that year. This means their figures include some deaths that occurred in years prior to the reference year (approximately 4%). As SID-Cymru will link and review data in the period leading up to death it is important to ensure the match/end date reflects the correct time period (specific to each individual case), i.e. date of death not date of registration, to afford an accurate perspective on utilisation of resources and help seeking behaviours. Consequently the actual date of death (ADOD) will be used in the matching criteria to establish a data review 'end date' for controls, rather than the registered date of death (RDOD) referred to in ONS reports. Mortality data within the SAIL Databank is only available from 2003, thus the earliest case inclusion relates to ADOD's from the 1st January 2003. To minimise the underestimation of cases identified (i.e., due to delays in COD confirmation and registration of DOD) the case inclusion for SID-Cymru includes cases aged 10 years and over at ADOD, where the ADOD took place between 01/01/2003 and 31/12/2011.

Table 1. ICD-10 codes used to identify and extract cases with death through probable suicide for SID-Cymru.

Short name	**Description**	**ICD10 codes**
Poisoning	Poisoning by and exposure to one or more types of poisons	X60-X69, Y10-Y19
Hanging & suffocation	Hanging, strangulation and suffocation	X70, Y20
Drowning	Drowning and submersion	X71, Y21
Guns	Firearm discharge	X72-X74, Y22-Y24
Sharp object	Sharp object	X78, Y28
Jumping from a high place	Jumping from a high place	X80, Y30
Jumping or lying before moving object	Jumping or lying before moving object	X81, Y31
Explosives & Fire	Explosive material, smoke, fire, flames and hot objects	X75-X77, Y25-Y27
Other or unspecified events	Intentional self-harm by other/un- specified means; Other/un- specified events undetermined intent	X83-X84, Y33-Y34 (excl.Y33.9)
Crash by motor vehicle	Crashing of motor vehicle intentional and undetermined intent	X82, Y32
Blunt Object	Blunt Object	X79, Y29

Table 2 presents the number of registered deaths through suicide per year and actual deaths per year between 2003 and 2011 as reported by the ONS and as identified for SID-Cymru within the SAIL Databank.

Table 2. Probable Suicide[1] Deaths for Wales 2003 – 2011 identified within the SAIL Databank for SID-Cymru.

Count by Year[2]	2003	2004	2005	2006	2007	2008	2009	2010	2011	TOTAL
Probable Suicides ADOD[3] within SID[4]	361	303	287	278	279	282	272	292	304	**2664**
Probable Suicides RDOD[5] within SID	274	330	270	300	288	266	257	287	340	**2612**
ONS Probable Suicides[6] RDOD	332	330	269	300	289	266	258	288	341	**2673**

[1]Probable suicides = ICD-10 codes for Intentional [X60-84], Undetermined [Y10-34 excl.Y33.9] self-harm; [2]Values are number of individuals unless otherwise indicated [3]ADOD = Year death occurred; [4]Cases identified for SID-Cymru includes those aged 10 years and over; [5]RDOD = Year death was registered; [6]Cases reported by the ONS include those aged 15 years and over.

Matched Controls will be identified within the WDS as live individuals matched on age (to the nearest year) and gender who were registered within the WDS for at least one year prior to matched case's ADOD. The controls will be required to be alive at the time of the matched case's ADOD (i.e., the match/end date). The use of live controls limits the introduction of bias relating to deaths, particularly in the younger age groups, of those undertaking risky behaviours resulting in premature death (e.g. substance use, accidents) that may be associated with unrecognised suicidal behaviours or known risk factors. To add to the power of the study we aim to identify at least five controls to every one case.[34]

Routine Data Sources

For SID-Cymru the data collected on identified cases and controls, via ADDE and the WDS respectively, will be linked to other routinely collected datasets, retrospectively allowing a review of each individual's pathway through the various services. Linkage with, for example, GP system data provides varying information about patients going back several years, including previous diagnosis, presenting symptoms and previous medications prescribed. This dataset can be used to review contacts with the GP and, consequently, infer the development / diagnosis of any new medical conditions including depression and self harm prior to suicide. Linkage with in-patient data will allow a review of hospital contacts and Emergency Department datasets will give information on crisis contacts. This will provide comprehensive insights into help seeking behaviours and management across settings. Data sets currently accessible via the SAIL Databank which will be linked to SID-Cymru are presented in Table 3. There are several other datasets currently under negotiation for inclusion within the SAIL Databank to which, in the future, SID-Cymru could link. These include Department of Work and Pensions (DWP) employment and incapacity status, Looked After Children, fostering, Substance Misuse Services, Sexually Transmitted Infections and Police Data.

Table 3. Datasets available within the SAIL Databank for linkage with SID-Cymru cases and controls.

Dataset	**Data Description**	**Data Provider**	**Data Type**	**Geographical Coverage**	**Available from**
Annual District Death Extract (ADDE)	Register of all deaths relating to Welsh residents, including those that died out of Wales.	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW) on behalf of ONS	Registry data	Wales	2003
Welsh Demographic Service (WDS)	Administrative information about individuals in Wales that use NHS services; such as address and practice registration history. It replaced the NHS Wales Administrative Register (NHSAR) in 2009.	NHS Wales' Informatics Service (NWIS)/ HSW	Administrative (mandatory) data return	Wales	1990
Area Based Deprivation Scores	Lower Super Output Area - Townsend Deprivation Index and Welsh Index of Multiple Deprivation Scores.	NHS Wales' Informatics Service (NWIS)/ HSW	Administrative data	Wales	2001 / 2005
Education Attainment (DCEL)	Educational attainment at key stages (pre 16 year old). Data collated from school census, the national pupil survey and educational attainment databases by the Welsh Government.	Department for Children, Education, Lifelong Learning and Skills (DCELLS)	Administrative / survey data.	Wales	2005
Primary Care GP dataset	Each GP practice uses a clinical information system to maintain an electronic health record for each of their patients; capturing the signs, symptoms, test results, diagnoses, prescribed treatment, referrals for specialist treatment and social aspects relating to the patients home environment.	Welsh General Practices that have signed up to the SAIL Databank	Clinical system data	Wales	2000
Patient Episode Database for Wales (PEDW)	NHS Wales - hospital admissions (Inpatients and day cases) dataset comprising of attendance and clinical information for all hospital admissions: includes diagnoses and operations performed.	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW)	Administrative (mandatory) data return	Wales	1997
Outpatient Dataset (OPD)	Attendance information for all NHS Wales hospital outpatient appointments.	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW)	Administrative (mandatory) data return	Wales	2004
Emergency Department Data Set (EDDS)	Administrative and clinical information for all NHS Wales Accident and Emergency department attendances. Includes the All Wales Injury Surveillance Systems (AWISS) dataset	NHS Wales' Informatics Service (NWIS)/ Health Solutions Wales (HSW)	Administrative (mandatory) data return	Wales	2009

Measurements

Data Variables/Characteristics

Data to be included in SID-Cymru will be extracted from the SAIL Databank and will include basic demographics; educational data; ADOD and RDOD; numbers and percentages for deprivation; proportions known to different healthcare settings in the period prior to death; inclusive medical history, i.e., primary care contact and diagnosis (by Read Codes), information about hospital / psychiatric admission and diagnosis (by ICD-10 Codes), and nature of service contact e.g. for self-harm, substance misuse. Therefore, the routinely collected data held by the NHS and other public bodies supplying the existing SAIL Databank will maximise the narrative of a death through suicide while being less resource intensive than psychological autopsies. The initial variables to be extracted and linked across data sets are described in Table 4.

Planned Analyses

The primary objective is to establish SID-Cymru as a resource for future analysis. It is important to maximise the utility of the resource, and some general principles of analysis have been determined.

Descriptive Epidemiology

Phase 1:

- Identification and description of Cases: Number of deaths registered with ONS and available within the SAIL Databank, with relevant ICD-10 codes defined as suicides, between 2003-2011.

Table 4. Characteristics to be identified and collected through data linkage for SID-Cymru Cases and Controls.

Data	**Characteristics**
Demographics	Sex
	Date of Birth
	Marital status*
	Employment*
	Education*
Mortality	Date of Death,
	Age at DOD (in years)
	Registered Date of Death
	Cause of Death (ICD-10 code)
Deprivation	LSOA
	WIMD
	Townsend
	Rural-Urban Indicators
Medical History	GP register*
	GP diagnosis (Read code)*
	GP events*
	Hospital attendance
	Hospital diagnosis
	Hospital events
	Outpatient attendance
	Outpatient diagnosis*
	Outpatient events
	ED attendance*
	ED diagnosis*
	ED events*
	Psychiatric Admission*

* Where data are available; there may be limitations to the number of cases with this information.

- Identification and description of Matched Controls.
- Basic demographics.
- Table of delay (days) in registering suicides and undetermined deaths.

Phase 2:

- Proportions known to different healthcare settings: number and percentages, with main diagnosis, that had a general hospital admission; emergency department contact for self-harm and other indications; psychiatric admission and primary care contact in the year prior to probable suicide.
- Numbers and percentages for deprivation, employment status, educational achievement and medical history (e.g., chronic pain, terminal illness, medication, previous self harm, and substance misuse) will also be sought.
- Number of cases with missing data across data sets for variables of interest will be noted.

Area based measures of socio-economic deprivation

Deprivation will be measured at Lower Super Output Area (LSOA) level using the Welsh Index of Multiple Deprivation[35] (WIMD) and Townsend Index Score.[36] All suicides and matched controls will be assigned to a lower layer super output area (LSOA). There are 1909 LSOAs in Wales with an average population of 1500 people (range: 1000-3000).[37] Linkage to WIMD and Townsend Index information is available in the SAIL Databank. These will be ranked for deprivation, divided into quintiles and standardised rates calculated.

Odds ratios for the described exposures in the case control study

A case-control study utilising SID-Cymru will be population based and so the relative risk of suicide will be estimated by conditional logistic regression model with SPSS (version 20). Crude odds ratios will be adjusted for general practice and/or LSOA by matching cases and controls. Unadjusted estimates, confounder-adjusted estimates and their precision (e.g., 95% confidence interval) will be produced. Interactions between variables will be assessed with the log likelihood ratio test based on results from the adjusted analysis. The population attributable risk will be calculated[38] on the basis of adjusted relative risks from the full analysis and the distribution of exposures in the cases.

We will also report information on the completeness of linkage with each dataset.

Ethics and Dissemination

Ethics

A large amount of preliminary work on anonymisation methodologies was undertaken to create the SAIL Databank system,[17- 18] and the SAIL Databank has the required ethical permissions and processes in place to analyse anonymised data. It operates within a robust series of guidelines in line with the Caldicott principles and the National Information Governance Board for Health and Social Care.[18] Ethical approval has been granted for SID-Cymru from HIRU's Information Governance Review Panel (IGRP), an independent body consisting of a range of government, regulatory and professional agencies. In compliance with the IGRP rulings and the Data Protection Act 1998[39] individual level data and personal identifier linkage codes will not be removed from the SAIL Databank and all analyses will be carried out within the SAIL Databank gateway at Swansea University, a secure access point to data within the SAIL Databank. The key points of the MRC/Wellcome Trust data sharing policy will be followed.[40 41]

Dissemination

This paper describes the protocol for the development of a Suicide Information Database (SID-Cymru), and the research opportunities available from an electronic case-control study of suicides within a whole population. SID-Cymru will have the ability to link suicide cases anonymously to primary and secondary health information along with other social care data, allowing us to review each case's journey through these datasets. The establishment of SID-Cymru and exploration of the linkage methodologies will improve our understanding of those who complete suicide (particularly those not known to mental health services) and will be used to inform service planning and policy decision making and implementation. It will help identify key opportunities and settings for prevention of this tragic event. By so doing, SID-Cymru will join other international databases of suicide research and provide a platform for further investigation and data linkages.

In order for SID-Cymru to become a functional resource it is important to be aware of the limits of health data available; though widely used in research, and offering a broad range of information about treatment and associated conditions, there are issues relating to determining the quality of patient records, the completeness of data available and any conclusions that may be drawn from them, perhaps particularly concerning primary care records.[42] That is, working with routinely collected data presented in its "raw" format requires a certain level of database analysis skills where duplicates, missing and erroneous entries are common occurrences. Whilst some such administrative / system-based recording issues are easy to identify and account for in individual datasets, it is not always apparent what is correct and what is erroneous at the combined level. Indeed, this problem is confounded when linkage of data reveals conflicting information causing routine data to appear inaccessible and attempts at linkage discouraged. Thus, a secondary aim for SID-Cymru is to share the skills developed as part of establishing a suicide database, which can

aid colleagues who may lack such analytical expertise and foster greater multi-disciplinary collaborations and advance suicide research.

The UK has a strong presence in form of a wide range of publications and expertise relating to suicide research. Successful and dedicated Suicide Research Centres exist in Bristol, Manchester and Oxford,[43-45] and Scotland recently commenced work on a Scottish Suicide Information Database ('ScotSID').[46] These centres of excellence report broadly about suicide, though for logistical reasons, often a regional focus is retained, relating to project funding, analytical and interventional work; for that reason the establishment of a Suicide Information Database in Wales will build on and enhance existing UK suicide resources and infrastructure since the level of linkage available within the SAIL Databank is unique in this field. Unique opportunities within the SAIL Databank include linkage across primary, secondary and emergency department data and with education data. Findings will be disseminated through publications in peer reviewed journals and presentations at local, national and international conferences. Communication and consultation with key stakeholders from health and social care (for example primary care, mental health, Royal Colleges), government and other policy makers, as well as, the third sector will occur. Dissemination will be facilitated by the wider roles and responsibilities in suicide prevention, nationally and internationally, of members of the research team.

Formatted: Font: Times New Roman, 12 pt

Implications and Significance

Valuable opportunities exist for a wide range of epidemiological and clinical studies on suicide in Wales and SID-Cymru has the potential to become an important resource in facilitating such research which will be of relevance internationally. In addition to the records that have already been included for linkage with SID-Cymru it is expected that over time relevant information from other data sources (e.g. the DWP) will be linked to SID-Cymru to provide a wider range of information on issues such as individuals' circumstances, the nature of their deaths, and their contact with extended services. Additionally, linkage can be made with non-routinely collected datasets such as that held by NCISH. Specific hypotheses that will be explored include: recency of primary care, hospital and emergency department contacts including attendance for self-harm, primary care diagnosis of depression, levels of treatment with anti-depressants and trends in such treatment over time, rural and urban geography, contacts for the elderly and levels of physical illness.

Findings of current and projected public health importance will be assessed and presented to support policy makers, commissioners and providers of health and social care in Wales. Non-identifiable information from this project will be made available to researchers in Wales, the UK and international collaborators. The initial project focus will be on identification of cases and controls, data linkage opportunities and methodological issues relating to the establishment of a Suicide Information Database and routine data linkage (Phase 1), before commencing data extraction and analysis (Phase 2). The proposed data collation and linkage of primary, secondary and emergency department health information together with educational data are currently unique in the United Kingdom. Thus~~Due to existing routine data linkages within the SAIL Databank~~[17] it is possible to develop a central repository for information relating to suicide in whole population that will be of relevance internationally. This paper describes the design and development of a Suicide Information Database in its infancy.

Authors' contributions

AJ, the principal investigator, was responsible for the conceptualization of SID-Cymru. AJ, KL, MD, DG, JS for the design of SID-Cymru; AJ and LK for its on-going operationalisation. AJ and LK drafted the manuscript. All authors read and approved the final manuscript.

Funding Statement

This study was funded by a grant from the National Institute for Social Care and Health Research, Welsh Government, grant number RFS-12-14. DG is a National Institute for Health Research (England) Senior Investigator.

Competing interests

The authors declare that they have no competing interests.

Ethics approval

Ethical approval has been granted by the Health Information Research Unit Information Governance Review Panel at the College of Medicine at Swansea University.

Data sharing statement

It will be possible to access the data after the publication of the results. Researchers interested in collaborations or further information are invited to contact AJ at a.john@swansea.ac.uk.

References

1. World Health Organisation. 66th World Health Assembly. Draft comprehensive mental health action plan 2013-2020. A66/10 Rev.1, WHO 16 May 2013. http://apps.who.int/gb/e/e_wha66.html Accessed September 2014.

2. Linsley KR, Schapira K, Kelly TP. Open verdict v. suicide—importance to research. *Br J Psychiatry* 2001, **178**:465-468.

3. Appleby L, Kapur N, Shaw J, *et al*. Annual Report: England, Northern Ireland, Scotland and Wales. Manchester: University of Manchester 2013. http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/AnnualReport2013_UK.pdf Accessed July 2014.

4. Andersen UA, Andersen M, Rosholm JU, *et al*. Contacts to the health care system prior to suicide: a comprehensive analysis using registers for general and psychiatric hospital admissions, contacts to general practitioners and practising specialists and drug prescriptions. *Acta Psychiatr Scand* 2000, **102**:126-34.

5. Qin P, Agerbo E, Mortensen PB. Suicide risk in relation to socioeconomic, demographic, psychiatric, and familial factors: a national register–based study of all suicides in Denmark, 1981–1997. *Am J Psychiatry* 2003, **160**:765-772.

6. Gunnell D, Löfving S, Gustafsson JE, *et al*. School performance and risk of suicide in early adulthood: follow-up of two national cohorts of Swedish schoolchildren. *Journal of affective disorders*, 2011, **131**:104-112.

7. Qin P, Hawton K, Mortensen PB, *et al*. Combined effects of physical illness and comorbid psychiatric disorder on risk of suicide in a national population study. *Br J Psychiatry* 2014, **204**:430-435.

8. Christiansen E, Larsen KJ, Agerbo E, *et al*. Incidence and risk factors for suicide attempts in a general population of young people: a Danish register-based study. *Aust N Zealand J Psychiatry* 2012, **47**: 259-270.

9. Björkenstam C, Johansson LA, Nordström P, *et al*. Suicide or undetermined intent? A register-based study of signs of misclassification. *Population health metrics*, 2014, **12**:11.

10. Hjorthøj CR, Madsen T, Agerbo E, *et al*. Risk of suicide according to level of psychiatric treatment: a nationwide nested case–control study. *Social psychiatry and psychiatric epidemiology*, 2014; 1-9.

11. Lawrence DM, Holman CD, Jablensky AV, *et al*. Suicide rates in psychiatric in-patients: an application of record linkage to mental health research**.** *Aust NZ J Public Health* 1999, **23**:468–470.

12. Lawrence D, D'Arcy C, Holman J, *et al*. Increasing rates of suicide in Western Australian psychiatric patients: a record linkage study. *Acta Psychiatr Scand* 2001, **104**:443-451.

13. Morrison KB, Laing L. *Adults' use of health services in the year before death by suicide in Alberta.* Statistics Canada; 2011.

14. Appleby L, Kapur N, Shaw J, *et al*. Suicide in primary care in England: 2002-2011. National Confidential Inquiry into Suicide and Homicide by People with Mental Illness (NCISH). Manchester: University of Manchester 2014. http://www.bbmh.manchester.ac.uk/cmhr/research/centreforsuicideprevention/nci/reports/SuicideinPrimaryCare2014.pdf Accessed August 2014.

15. ISD Scotland: The Scottish Suicide Information Database Report 2014. http://www.isdscotland.org/Health-Topics/Public-Health/Publications/index.asp?ID=1232 Accessed August 2014.

16. Webb RT, Kontopantelis E, Doran T, *et al*. Suicide risk in primary care patients with major physical diseases: a case-control study. *Arch Gen Psychiatry* 2012, **69**:256-264.

17. Lyons RA, Jones KH, John G, *et al*. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Making* 2009, **9**:3.

18. Ford DV, Jones KH, Verplancke JP, *et al*. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009, **9**:1.

19. World Health Organization: *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. WHO: Geneva; 1993.

20. Office of National Statistics. 2011 Census Results for Wales. ONS 16th July 2012. http://wales.gov.uk/docs/statistics/2012/120716-census-2011-first-results-wales-en.pdf Accessed September 2014.

21. Office of National Statistics. Statistical bulletin: Suicide rates in the United Kingdom, 2006 to 2010. ONS 26th January 2012. http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2010/stb-statistical-bulletin.html Accessed August 2014.

22. McGregor JI, Brooks CJ, Chalasani P, *et al*. The Health Informatics Trial Enhancement Project (HITE): Using routinely collected primary care data to identify potential participants for a depression trial. *Trials* 2010, **11**:39.

23. Brooks CJ, Stephens JW, Price DE, *et al*. Use of a patient linked data warehouse to facilitate diabetes trial recruitment from primary care. *Primary Care Diabetes* 2009, **3**:245-248

24. Hill RA, Brophy S, Brunt H, *et al*. Protocol of the baseline assessment for the Environments for Healthy Living (EHL) Wales cohort study. *BMC Public Health* 2010, **10**:150

25. Fone DL, Dunstan F, White J, *et al*. Cohort Profile: The Caerphilly Health and Social Needs Electronic Cohort Study (E-CATALyST). *International Journal of Epidemiology*, 2013, **42**:1620-1628.

26. Rodgers SE, Heaven M, Lacey A, et al. Cohort profile: the housing regeneration and health study. *Int J Epidemiol* 2012, dys200.

27. SAIL Databank: Core Data Sets. Swansea University: Swansea http://www.saildatabank.com/data-dictionary/sail-datasets Accessed September 2014.

28. Brock A, Baker A, Griffiths C, *et al*. Suicide trends and geographical variations in the United Kingdom, 1991–2004. *Health Statistics Quarterly* 2006, **31**:6-22.

29. Carroll R, Hawton K, Kapur N, *et al*. Impact of the growing use of narrative verdicts by coroners on geographic variations in suicide: analysis of coroners' inquest data. *Journal of Public Health* 2011, **10**:1-7.

30. Gunnell D, Hawton, K, Kapur N. Coroners' verdicts and suicide statistics in England and Wales. *BMJ* 2011, **343**:d6030.

31. Hill C, Cook, L. Narrative verdicts and their impact on mortality statistics in England and Wales. *Health Statistics Quarterly* 2011, **49**:81-100.

32. Gunnell D, Bennewith O, Simkin S, *et al*. Time trends in coroners' use of different verdicts for possible suicides and their impact on officially reported incidence of suicide in England: 1990-2005. *Psychol Med* 2013, **43**:1415-1422.

33. Office of National Statistics. Statistical bulletin: Suicide rates in the United Kingdom, 2011. ONS 22 January 2013. http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2011/stb-suicide-bulletin.html Accessed August 2014.

34. Grimes DA, Schultz KF. Compared to what? Finding controls in a case control study. *Lancet* 2005, **365**: 1429–33.

35. Office for National Statistics. Welsh index of Multiple Deprivation 2011. Summary report, Welsh Government; 2011.

36. Townsend P, Phillimore P, Beattie A. *Health and Deprivation: Inequality and the North*. Routledge: London; 1988.

37. Office for National Statistics. Super Output Areas 2011. http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html Accessed August 2014.

38. Bruzzi P, Green SB, Byar DP, *et al*. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol* 1985, **122**:904-914.

39. Great Britain: Data Protection Act 1998. London: HMSO. http://www.legislation.gov.uk/ukpga/1998/29/contents Accessed August 2014.

40. MRC: Policy and Guidance on Sharing of Research Data from Population and Patient Studies. MRC 23rd November 2011. http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC008302 Accessed August 2014.

41. Wellcome Trust: Data sharing policy. Wellcome Trust August 2010. http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm Accessed August 2014.

42. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ* 2003, **326**:1070.

43. University of Bristol, Suicide Research Centre. http://www.bristol.ac.uk/psychiatry/research/suicideresearch.html Accessed August 2014.

44. The University of Manchester Centre for Suicide Prevention. http://www.bbmh.manchester.ac.uk/cmhr/research/centreforsuicideprevention/ Accessed August 2014.

45. University of Oxford, Centre for Suicide Research. http://cebmh.warne.ox.ac.uk/csr/ Accessed August 2014.

46. ISD Scotland. Scottish Suicide Information Database (ScotSID). ISD Scotland 2011. http://data.gov.uk/dataset/scottish_suicide_information_database Accessed August 2014.

STROBE Statement—checklist of items that should be included in reports of observational studies-protocol limits some of these

	Item No	Recommendation	Line
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	33
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	34-41
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	73-131
Objectives	3	State specific objectives, including any prespecified hypotheses	136-150
Methods			
Study design	4	Present key elements of study design early in the paper	152-7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	158-300
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	209-282
		(*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	237-282
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Table 3 , table 4, 283-349
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	Table 3 , table 4, 283-349
Bias	9	Describe any efforts to address potential sources of bias	350-360
Study size	10	Explain how the study size was arrived at	245-282, table 2
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	314-359:protocol
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	314-359:protocol
		(*b*) Describe any methods used to examine subgroups and interactions	314-359:protocol
		(*c*) Explain how missing data were addressed	314-359:protocol
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed	236-282

		Case-control study—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy
		(e) Describe any sensitivity analyses

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Table 2, protocol
		(b) Give reasons for non-participation at each stage	Will
		(c) Consider use of a flow diagram	Will
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Will
		(b) Indicate number of participants with missing data for each variable of interest	Will
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
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	Will
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Will
		(*b*) Report category boundaries when continuous variables were categorized	Will
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Will
Discussion			
Key results	18	Summarise key results with reference to study objectives	Will
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	Will
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Will
Generalisability	21	Discuss the generalisability (external validity) of the study results	Will
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	447

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.