

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 **Suicide Information Database- Cymru: A protocol for a population based, routinely**
4 **collected data linkage study to explore risks and patterns of healthcare contact prior to**
5 **suicide to identify opportunities for intervention.**
6
7
8

9
10 John A^{1*}, Dennis M¹, Kosnes L¹, Gunnell D², Scourfield J³, Ford DV¹, Lloyd K¹
11

12
13 ¹College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park,
14 Swansea, SA2 8PP, United Kingdom.
15

16
17
18 ²School of Social and Community Medicine, University of Bristol, Canynge Hall, 39
19 Whatley Road, Bristol, BS8 2PS, United Kingdom.
20

21
22
23 ³School of Social Sciences, Cardiff University, Glamorgan Building, King Edward VII
24 Avenue, Cardiff, CF10 3WT, United Kingdom.
25
26
27

28
29
30
31
32 *Corresponding Author: Ann John^{1*} a.john@swansea.ac.uk; Phone: +44(0)1792 602568;
33
34 College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park,
35 Swansea, SA2 8PP, United Kingdom.
36
37

38
39 Michael Dennis¹ m.s.dennis@swansea.ac.uk, Liv Kosnes¹ l.kosnes@swansea.ac.uk, David
40 Gunnell² d.j.gunnell@bristol.ac.uk, Jonathan Scourfield³ Scourfield@cardiff.ac.uk, David V.
41 Ford¹ d.v.ford@swansea.ac.uk, Keith Lloyd¹ k.r.lloyd@swansea.ac.uk
42
43
44
45
46

47 **Key words:** Suicide, case-control, database, routine data linkage.
48

49
50 **Word Count: 4054**
51
52
53
54
55
56
57
58
59
60

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

1
2
3 improved understanding of the prior health, nature of previous contacts with services, and
4
5 wider social circumstances of those who complete suicide will assist in prevention policy,
6
7 service organisation and delivery. SID-Cymru is funded through the National Institute for
8
9 Social Care and Health Research, Welsh Government (RFS-12-14).
10

11 12 13 **Article Summary**

14 15 16 Article Focus:

- 17
18 • Case definitions of probable and possible suicides
- 19
20
21 • Identifying risk factors and patterns of service use in those who complete suicide
- 22
23
24
25 • Designing a study
- 26

27 28 Key messages:

- 29
30
31 • SID-Cymru will facilitate a series of case- control studies of those who complete
32
33 suicide
- 34
35
36 • Recency of primary care, in- and out-patient hospital and emergency department
37
38 contacts including attendance for self-harm, primary care diagnosis of depression,
39
40 levels of treatment with anti-depressants and trends in such treatment over time, rural
41
42 and urban geography, educational attainment, contacts for the elderly and levels of
43
44 physical illness will be investigated
- 45
46
47 • Results will be adjusted for a number of confounders
- 48

49 50 Strengths and Limitations of this study:

- 51
52
53 • Based on the general population
- 54
55
56 • Includes all deaths through suicide, not just those known to mental health services
- 57

- 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

For peer review only

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 essential¹. 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 Others¹¹⁻¹³ 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.

1
2
3 In the United Kingdom various systems exist to examine suicide deaths. The National
4 Confidential Inquiry into Suicide and Homicide (NCISH) by people with Mental Illness
5 focuses on suicide cases that were in contact with mental health services (approximately 25%
6 of total) in the year prior to their deaths.³ However this provides limited information on
7 issues of suicide in the general population and may hamper the effectiveness of wider
8 preventive efforts.¹ In a recent report NCISH¹⁴ examined aspects of primary healthcare prior
9 to all suicides in England between 2002-2011; however no linkage was made with data from
10 other service providers such as emergency departments.
11
12
13
14
15
16
17
18
19

20
21 Scotland has recently established enhanced data collection in relation to suicide, however,
22 further development is needed in order for 'ScotSID' to be able to examine healthcare
23 pathways and contact with more than one health service;¹⁵ England does not currently have a
24 dedicated repository for suicide data though studies have made use of the Clinical Practice
25 Research Datalink (CPRD; www.cprd.com), which represents approximately 8.5% of the UK
26 population from 600 General Practices (GP's) in England.¹⁶ CPRD can be linked with data
27 from the National Health Service (NHS) Hospital Episode Statistics (HES;
28 <http://www.hscic.gov.uk>) and mortality data from the Office of National Statistics (ONS;
29 <http://www.ons.gov.uk>) but has limited emergency department data.
30
31
32
33
34
35
36
37
38
39
40
41

42 SID-Cymru will access and link information on prior health, nature of previous contacts with
43 services, and wider social circumstances of all those who complete suicide (known and
44 unknown to mental health services) within the population of Wales via anonymised routinely
45 collected electronic data held in healthcare and social datasets from the Secure Anonymised
46 Information Linkage (SAIL) Databank (www.saildatabank.com).^{17 18} This brings together,
47 links and anonymises the widest possible range of person-based data currently available in
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 the UK. The SAIL Databank was originally set up by the Health Information Research Unit
4
5 (HIRU) at the College of Medicine at Swansea University.
6
7

8
9 SID-Cymru is part of the research programme related to the Health e-Research Collaboration
10
11 UK (HeRC UK), led by the Medical Research Council (MRC) and based in the Centre for the
12
13 Improvement of Population Health through e-Records Research (CIPHER). CIPHER is a UK
14
15 Clinical Research Collaboration (UKCRC) Public Health Research Centre of Excellence set
16
17 within the College of Medicine at Swansea University. SID-Cymru will provide timely robust
18
19 data to inform the future strategic direction of and the first step in designing and evaluating
20
21 effective interventions to prevent suicide. It will use the International Classification of
22
23 Disease, Tenth Revision,¹⁹ (ICD-10) definitions and instructions for classifying causes of
24
25 death which will allow for comparisons with other countries and thus support research of
26
27 relevance globally. An emphasis will be given to issues where there are opportunities for
28
29 intervention or where electronic data linkage confers an advantage to its investigation.
30
31

32 33 **Aim**

34
35
36 The main aim is to establish a Suicide Information Database- Cymru as a population based
37
38 resource for studying factors and service contacts associated with all suicide deaths through
39
40 routinely collected data linkage case control studies.
41
42

43 44 **Objectives:**

45 46 *Phase 1:*

- 47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
 - 50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

- 1
2
3
4
5
6
7
8
9
- 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.

10
11

Phase 2:

- 12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
- 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.

27
28

Methods and Analyses

29
30
31

Design

32
33
34
35
36
37
38
39
40
41
42
43
44

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 million²⁰ 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

45
46
47

Data Source

48
49
50
51
52
53
54
55
56

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

1
2
3 Within the SAIL Databank a split-file approach to anonymisation is used to overcome issues
4 of confidentiality and disclosure in health-related data warehousing.¹⁸ Demographic data are
5 sent to a partner organisation, NHS Wales Information Service (NWIS), where all identifiable
6 information is removed; clinical data are sent directly to HIRU, where, for each dataset
7 within the SAIL Databank, an individual is assigned an encrypted Anonymised Linking Field
8 (ALF). The ALF is used to link anonymised individuals across datasets, thus supporting the
9 opportunity to conduct longitudinal analyses of an individual's journey through multiple
10 health, education and social datasets.¹⁷ Additionally Residential Anonymised Linking Fields
11 (RALFs) have been created for all residences in Wales and enables linkage of anonymised
12 household and environment data with the health records of individual residents without the
13 identity of the residences or residents being known to researchers.²⁶

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28 *The primary study base* will be the Welsh Demographic Service²⁷ (WDS) hosted within the
29 SAIL Databank. The WDS is a core dataset available within the SAIL Databank and part of a
30 set of services to manage administrative information (demographic data) for NHS patients in
31 Wales. The WDS was introduced early in 2009 replacing a similar service known as the NHS
32 Wales Administrative Register (NHS AR). The WDS data is collected from GP practices via
33 the Exeter System and more than five million individuals are currently present in the WDS
34 dataset within the SAIL Databank. The WDS is a register of all individuals that have at some
35 point in time been registered with a Welsh General Practitioner (GP) or required some form
36 of NHS health care provision in Wales. The electronic collation of WDS / NHS AR data
37 originated in 1960 and is updated and maintained by NWIS,²⁷ ensuring that address changes
38 (within and out of Wales) and death notices are included in the register. The original (non-
39 anonymised) version of the NHS AR has been used in the HIRU Matching Algorithm for
40 Consistent Results in Anonymised Linkage (MACRAL), making the WDS / NHS AR the
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 master list for all Welsh residents and using probabilistic matching to find the associated
4
5 NHS numbers that are then encrypted into ALF's.
6
7

8 Deaths in Wales should be registered within five days of the date of death. However
9
10 legislation in Wales means that when a coroner's inquest takes place, the death cannot be
11
12 registered until the inquest is complete. Since the Office for National Statistics (ONS), the
13
14 national agency where all deaths are collated, has no conclusive information about the death
15
16 until it is registered, there is a delay between the date the death occurred and when the death
17
18 is added to the annual ONS mortality dataset. The ONS found that less than 41% of deaths
19
20 going to inquest were registered within 3 months, though 96% were registered within 1
21
22 year.²⁸ Thus precise information on the annual incidence of suicide may be delayed by up to
23
24 two years. Information collected at death registration is recorded on the Registration Online
25
26 system by registrars. Most of the information is normally supplied by the informant (usually a
27
28 close relative of the deceased) while the cause of death (COD) is usually obtained from the
29
30 Medical Certificate of Cause of Death (MCCD) completed by a medical practitioner when
31
32 the death is certified using ICD-10 coding or the coroner if there is an inquest.¹⁹ Notably, a
33
34 death is not officially registered within the Annual District Deaths Extract (ADDE) until the
35
36 COD has been finalised, and thus the year of death and the year of registration may not
37
38 concur. The *primary dataset* used to construct SID-Cymru is the ADDE from the ONS. The
39
40 ADDE is inclusive of Welsh residents who died outside of Wales, and holds information
41
42 about COD derived from death certificates on all deaths in Wales.
43
44
45
46
47
48

49 *Definition of Suicide for Cases*

50
51

52 The true number of suicides is difficult to determine because a coroner's conclusion of
53
54 suicide must be 'beyond a reasonable doubt' i.e. that the death was intentionally self inflicted
55
56 and in some areas coroners have increasingly (since 2001) reported narrative conclusions
57
58

1
2
3 rather than reporting it as suicide.^{29 30} Previously when insufficient information was recorded
4
5 by the coroner ONS coders used to record the death as an accident which inevitably led to
6
7 some suicides being classified as accidents or misadventure. The ONS has recently issued
8
9 guidance on this issue following a coding practice review.³¹ Current ONS practice includes
10
11 deaths where intention is ‘undetermined whether accidentally or purposefully inflicted’; thus
12
13 deaths where there may be no intention to take life, such as in relation to injury or poisoning,
14
15 are included in suicide figures by ONS. Currently there is no access to coroners’ narrative
16
17 verdicts within the SAIL Databank as a possible method for review of case inclusion.
18
19

20
21
22 There is evidence to suggest that a high proportion of deaths from poisoning and hanging that
23
24 receive accidental verdicts, are found, when subjected to clinical review, to be suicides.³²
25
26 Such possible deaths through suicide will be included in SID-Cymru as an opportunity for
27
28 further separate and combined analysis; thus the additional ICD-10 codes relating to
29
30 ‘accidental poisoning with prescribed drugs’ (X40-X41, X43-X49) and ‘accidental hanging’
31
32 (W75-W76) may be used along with ‘sequelae of external causes of morbidity and mortality’
33
34 (Y87, Y87.2, Y89, Y89.9).¹⁸ Thus with SID-Cymru we aim to establish a Suicide
35
36 Information Database that can be a resource for studying factors associated with *all*
37
38 probable, as well as, possible deaths through suicide; by including accidental
39
40 hanging/strangulation and accidental poisoning, excluding, narcotics and psychodysleptics
41
42 (i.e., possible suicides), along with the ONS traditional method of defining suicides, which
43
44 includes suicides and deaths of undetermined intent (probable suicides).²
45
46
47

48 49 *Identification of Cases and Controls*

50
51
52 SID-Cymru cases encompass suicides (‘intentional self-harm’ [ICD-10: X60-84]) and
53
54 probable deaths through suicide (‘undetermined intent’ [ICD-10: Y10-34, excl.Y33.9])
55
56 recorded from the MCCD and presented within ADDE as the underlying COD. Those coded
57
58
59
60

1
2
3 Y33.9/U50.9 (pending verdicts) are excluded since a large proportion of these are
4
5 subsequently found to be homicides. Probable suicide defined above can be supplemented
6
7 with possible deaths through suicide if required. Cases of probable suicide will be identified
8
9 and extracted by use of the ICD-10 codes¹⁹ defined above and depicted in Table 1.
10
11

12
13 ONS figures include those over 15 years of age only, due to the possibility that deaths in
14
15 younger children coded as undetermined events may be caused by unverifiable accidents,
16
17 neglect or abuse.³³ SID-Cymru will allow for the analysis of suicides and probable suicides
18
19 for the 10 to 14 year age band. Official ONS mortality statistics are produced based on the
20
21 number of deaths registered in a particular calendar year, rather than the number of deaths
22
23 that occurred in that year. This means their figures include some deaths that occurred in years
24
25 prior to the reference year (approximately 4%). As SID-Cymru will link and review data in
26
27 the period leading up to death it is important to ensure the match/end date reflects the correct
28
29 time period (specific to each individual case), i.e. date of death not date of registration, to
30
31 afford an accurate perspective on utilisation of resources and help seeking behaviours.
32
33 Consequently the actual date of death (ADOD) will be used in the matching criteria to
34
35 establish a data review 'end date' for controls, rather than the registered date of death
36
37 (RDOD) referred to in ONS reports. Mortality data within the SAIL Databank is only
38
39 available from 2003, thus the earliest case inclusion relates to ADOD's from the 1st January
40
41 2003. To minimise the underestimation of cases identified (i.e., due to delays in COD
42
43 confirmation and registration of DOD) the case inclusion for SID-Cymru includes cases aged
44
45 10 years and over at ADOD, where the ADOD took place between 01/01/2003 and 31/12/
46
47 2011.
48
49
50
51
52
53
54
55
56
57
58
59
60

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 Suicide¹ 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 ADOD³ within SID⁴	361	303	287	278	279	282	272	292	304	2664
Probable Suicides RDOD⁵ within SID	274	330	270	300	288	266	257	287	340	2612
ONS Probable Suicides⁶ RDOD	332	330	269	300	289	266	258	288	341	2673

¹Probable suicides = ICD-10 codes for Intentional [X60-84], Undetermined [Y10-34 excl. Y33.9] self-harm;

²Values are number of individuals unless otherwise indicated ³ADOD = Year death occurred; ⁴Cases identified for SID-Cymru includes those aged 10 years and over; ⁵RDOD = Year death was registered; ⁶Cases 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 Deprivation³⁵ (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

1
2
3 A case-control study utilising SID-Cymru will be population based and so the relative risk of
4
5 suicide will be estimated by conditional logistic regression model with SPSS (version 20).
6
7 Crude odds ratios will be adjusted for general practice and/or LSOA by matching cases and
8
9 controls. Unadjusted estimates, confounder-adjusted estimates and their precision (e.g., 95%
10
11 confidence interval) will be produced. Interactions between variables will be assessed with
12
13 the log likelihood ratio test based on results from the adjusted analysis. The population
14
15 attributable risk will be calculated³⁸ on the basis of adjusted relative risks from the full
16
17 analysis and the distribution of exposures in the cases.
18
19

20
21 We will also report information on the completeness of linkage with each dataset.
22
23

24 **Ethics and Dissemination**

25 *Ethics*

26
27
28 A large amount of preliminary work on anonymisation methodologies was undertaken to
29
30 create the SAIL Databank system,^{17- 18} and the SAIL Databank has the required ethical
31
32 permissions and processes in place to analyse anonymised data. It operates within a robust
33
34 series of guidelines in line with the Caldicott principles and the National Information
35
36 Governance Board for Health and Social Care.¹⁸ Ethical approval has been granted for SID-
37
38 Cymru from HIRU's Information Governance Review Panel (IGRP), an independent body
39
40 consisting of a range of government, regulatory and professional agencies. In compliance
41
42 with the IGRP rulings and the Data Protection Act 1998³⁹ individual level data and personal
43
44 identifier linkage codes will not be removed from the SAIL Databank and all analyses will be
45
46 carried out within the SAIL Databank gateway at Swansea University, a secure access point
47
48 to data within the SAIL Databank. The key points of the MRC/Wellcome Trust data sharing
49
50 policy will be followed.^{40 41}
51
52
53
54
55
56
57
58
59
60

1
2
3 This paper describes the protocol for the development of a Suicide Information Database
4 (SID-Cymru), and the research opportunities available from an electronic case-control study
5 of suicides within a whole population. SID-Cymru will have the ability to link suicide cases
6 anonymously to primary and secondary health information along with other social care data,
7 allowing us to review each case's journey through these datasets. The establishment of SID-
8 Cymru and exploration of the linkage methodologies will improve our understanding of those
9 who complete suicide (particularly those not known to mental health services) and will be
10 used to inform service planning and policy decision making and implementation. It will help
11 identify key opportunities and settings for prevention of this tragic event. By so doing, SID-
12 Cymru will join other international databases of suicide research and provide a platform for
13 further investigation and data linkages.
14
15
16
17
18
19
20
21
22
23
24
25
26

27 In order for SID-Cymru to become a functional resource it is important to be aware of the
28 limits of health data available; though widely used in research, and offering a broad range of
29 information about treatment and associated conditions, there are issues relating to
30 determining the quality of patient records, the completeness of data available and any
31 conclusions that may be drawn from them, perhaps particularly concerning primary care
32 records.⁴² That is, working with routinely collected data presented in its "raw" format
33 requires a certain level of database analysis skills where duplicates, missing and erroneous
34 entries are common occurrences. Whilst some such administrative / system-based recording
35 issues are easy to identify and account for in individual datasets, it is not always apparent
36 what is correct and what is erroneous at the combined level. Indeed, this problem is
37 confounded when linkage of data reveals conflicting information causing routine data to
38 appear inaccessible and attempts at linkage discouraged. Thus, a secondary aim for SID-
39 Cymru is to share the skills developed as part of establishing a suicide database, which can
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 aid colleagues who may lack such analytical expertise and foster greater multi-disciplinary
4
5 collaborations and advance suicide research.
6
7

8 The UK has a strong presence in form of a wide range of publications and expertise relating
9
10 to suicide research. Successful and dedicated Suicide Research Centres exist in Bristol,
11
12 Manchester and Oxford,⁴³⁻⁴⁵ and Scotland recently commenced work on a Scottish Suicide
13
14 Information Database ('ScotSID').⁴⁶ These centres of excellence report broadly about suicide,
15
16 though for logistical reasons, often a regional focus is retained, relating to project funding,
17
18 analytical and interventional work; for that reason the establishment of a Suicide Information
19
20 Database in Wales will build on and enhance existing UK suicide resources and infrastructure
21
22 since the level of linkage available within the SAIL Databank is unique in this field. Unique
23
24 opportunities within the SAIL Databank include linkage across primary, secondary and
25
26 emergency department data and with education data.
27
28
29

30 31 **Implications and Significance** 32

33
34 Valuable opportunities exist for a wide range of epidemiological and clinical studies on
35
36 suicide in Wales and SID-Cymru has the potential to become an important resource in
37
38 facilitating such research which will be of relevance internationally. In addition to the records
39
40 that have already been included for linkage with SID-Cymru it is expected that over time
41
42 relevant information from other data sources (e.g. the DWP) will be linked to SID-Cymru to
43
44 provide a wider range of information on issues such as individuals' circumstances, the nature
45
46 of their deaths, and their contact with extended services. Additionally, linkage can be made
47
48 with non-routinely collected datasets such as that held by NCISH. Specific hypotheses that
49
50 will be explored include: recency of primary care, hospital and emergency department
51
52 contacts including attendance for self-harm, primary care diagnosis of depression, levels of
53
54
55
56
57
58
59
60

1
2
3 treatment with anti-depressants and trends in such treatment over time, rural and urban
4
5 geography, contacts for the elderly and levels of physical illness.
6

7 Findings of current and projected public health importance will be assessed and presented to
8
9 support policy makers, commissioners and providers of health and social care in Wales. Non-
10
11 identifiable information from this project will be made available to researchers in Wales, the
12
13 UK and international collaborators. The initial project focus will be on identification of cases
14
15 and controls, data linkage opportunities and methodological issues relating to the
16
17 establishment of a Suicide Information Database and routine data linkage (Phase 1), before
18
19 commencing data extraction and analysis (Phase 2). The proposed data collation and linkage
20
21 of primary, secondary and emergency department health information together with
22
23 educational data are currently unique in the United Kingdom. Due to existing routine data
24
25 linkages within the SAIL Databank¹⁷ it is possible to develop a central repository for
26
27 information relating to suicide in whole population that will be of relevance internationally.
28
29 This paper describes the design and development of a Suicide Information Database in its
30
31 infancy.
32
33
34
35

36 37 **Authors' contributions**

38
39
40 AJ, the principal investigator, was responsible for the conceptualization of SID-Cymru. AJ,
41
42 KL, MD, DG, JS for the design of SID-Cymru; AJ and LK for its on-going
43
44 operationalisation. AJ and LK drafted the manuscript. All authors read and approved the final
45
46 manuscript.
47

48 49 **Funding Statement**

1
2
3 This study was funded by a grant from the National Institute for Social Care and Health
4
5 Research, Welsh Government, grant number RFS-12-14. DG is a National Institute for
6
7 Health Research (England) Senior Investigator.
8
9

10 **Competing interests**

11
12 The authors declare that they have no competing interests.
13
14

15 **Ethics approval**

16
17 Ethical approval has been granted by the Health Information Research Unit Information
18
19 Governance Review Panel at the College of Medicine at Swansea University.
20
21
22
23

24 **Data sharing statement**

25
26 It will be possible to access the data after the publication of the results. Researchers interested
27
28 in collaborations or further information are invited to contact AJ at a.john@swansea.ac.uk.
29
30
31

32 **References**

- 33
34
35
36 1. World Health Organisation. 66th World Health Assembly. Draft comprehensive
37
38 mental health action plan 2013-2020. A66/10 Rev.1, WHO 16 May 2013.
39
40 http://apps.who.int/gb/e/e_wha66.html Accessed September 2014.
41
42
43 2. Linsley KR, Schapira K, Kelly TP. Open verdict v. suicide—importance to research.
44
45 *Br J Psychiatry* 2001, **178**:465-468.
46
47
48 3. Appleby L, Kapur N, Shaw J, *et al*. Annual Report: England, Northern Ireland,
49
50 Scotland and Wales. Manchester: University of Manchester 2013.
51
52 [http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/Ann](http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/AnnualReport2013_UK.pdf)
53
54 [ualReport2013_UK.pdf](http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/AnnualReport2013_UK.pdf) Accessed July 2014.
55
56
57
58
59
60

- 1
2
3 4. Andersen UA, Andersen M, Rosholm JU, *et al.* Contacts to the health care system
4
5 prior to suicide: a comprehensive analysis using registers for general and psychiatric
6
7 hospital admissions, contacts to general practitioners and practising specialists and
8
9 drug prescriptions. *Acta Psychiatr Scand* 2000, **102**:126-34.
- 10
11 5. Qin P, Agerbo E, Mortensen PB. Suicide risk in relation to socioeconomic,
12
13 demographic, psychiatric, and familial factors: a national register-based study of all
14
15 suicides in Denmark, 1981–1997. *Am J Psychiatry* 2003, **160**:765-772.
- 16
17 6. Gunnell D, Löfving S, Gustafsson JE, *et al.* School performance and risk of suicide in
18
19 early adulthood: follow-up of two national cohorts of Swedish schoolchildren.
20
21 *Journal of affective disorders*, 2011, **131**:104-112.
- 22
23 7. Qin P, Hawton K, Mortensen PB, *et al.* Combined effects of physical illness and
24
25 comorbid psychiatric disorder on risk of suicide in a national population study. *Br J*
26
27 *Psychiatry* 2014, **204**:430-435.
- 28
29 8. Christiansen E, Larsen KJ, Agerbo E, *et al.* Incidence and risk factors for suicide
30
31 attempts in a general population of young people: a Danish register-based study. *Aust*
32
33 *N Zealand J Psychiatry* 2012, **47**: 259-270.
- 34
35 9. Björkenstam C, Johansson LA, Nordström P, *et al.* Suicide or undetermined intent? A
36
37 register-based study of signs of misclassification. *Population health metrics*, 2014,
38
39 **12**:11.
- 40
41 10. Hjorthøj CR, Madsen T, Agerbo E, *et al.* Risk of suicide according to level of
42
43 psychiatric treatment: a nationwide nested case-control study. *Social psychiatry and*
44
45 *psychiatric epidemiology*, 2014; 1-9.

- 1
2
3 11. Lawrence DM, Holman CD, Jablensky AV, *et al.* Suicide rates in psychiatric in-
4
5 patients: an application of record linkage to mental health research. *Aust NZ J Public*
6
7 *Health* 1999, **23**:468–470.
- 8
9
10 12. Lawrence D, D'Arcy C, Holman J, *et al.* Increasing rates of suicide in Western
11
12 Australian psychiatric patients: a record linkage study. *Acta Psychiatr Scand* 2001,
13
14 **104**:443-451.
- 15
16
17 13. Morrison KB, Laing L. *Adults' use of health services in the year before death by*
18
19 *suicide in Alberta*. Statistics Canada; 2011.
- 20
21 14. Appleby L, Kapur N, Shaw J, *et al.* Suicide in primary care in England: 2002-2011.
22
23 National Confidential Inquiry into Suicide and Homicide by People with Mental
24
25 Illness (NCISH). Manchester: University of Manchester 2014.
26
27 <http://www.bbmh.manchester.ac.uk/cmhr/research/centreforsuicideprevention/nci/reports/SuicideinPrimaryCare2014.pdf> Accessed August 2014.
- 28
29
30
31
32 15. ISD Scotland: The Scottish Suicide Information Database Report 2014.
33
34 <http://www.isdscotland.org/Health-Topics/Public-Health/Publications/index.asp?ID=1232> Accessed August 2014.
- 35
36
37
38
39 16. Webb RT, Kontopantelis E, Doran T, *et al.* Suicide risk in primary care patients with
40
41 major physical diseases: a case-control study. *Arch Gen Psychiatry* 2012, **69**:256-264.
- 42
43
44 17. Lyons RA, Jones KH, John G, *et al.* The SAIL databank: linking multiple health and
45
46 social care datasets. *BMC Med Inform Decis Making* 2009, **9**:3.
- 47
48
49
50 18. Ford DV, Jones KH, Verplancke JP, *et al.* The SAIL Databank: building a national
51
52 architecture for e-health research and evaluation. *BMC Health Serv Res* 2009, **9**:1.
- 53
54
55
56
57
58
59
60

- 1
2
3 19. World Health Organization: *The ICD-10 classification of mental and behavioural*
4 *disorders: Clinical descriptions and diagnostic guidelines*. WHO: Geneva; 1993.
5
6
7
8 20. Office of National Statistics. 2011 Census Results for Wales. ONS 16th July 2012.
9
10 <http://wales.gov.uk/docs/statistics/2012/120716-census-2011-first-results-wales->
11 [en.pdf](#) Accessed September 2014.
12
13
14
15 21. Office of National Statistics. Statistical bulletin: Suicide rates in the United Kingdom,
16 2006 to 2010. ONS 26th January 2012. [http://www.ons.gov.uk/ons/rel/subnational-](http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2010/stb-statistical-bulletin.html)
17 [health4/suicides-in-the-united-kingdom/2010/stb-statistical-bulletin.html](#) Accessed
18 August 2014.
19
20
21
22
23
24
25 22. McGregor JJ, Brooks CJ, Chalasani P, *et al*. The Health Informatics Trial
26 Enhancement Project (HITE): Using routinely collected primary care data to identify
27 potential participants for a depression trial. *Trials* 2010, **11**:39.
28
29
30
31
32
33 23. Brooks CJ, Stephens JW, Price DE, *et al*. Use of a patient linked data warehouse to
34 facilitate diabetes trial recruitment from primary care. *Primary Care Diabetes* 2009,
35 **3**:245-248
36
37
38
39
40 24. Hill RA, Brophy S, Brunt H, *et al*. Protocol of the baseline assessment for the
41 Environments for Healthy Living (EHL) Wales cohort study. *BMC Public Health*
42 2010, **10**:150
43
44
45
46
47
48 25. Fone DL, Dunstan F, White J, *et al*. Cohort Profile: The Caerphilly Health and Social
49 Needs Electronic Cohort Study (E-CATALyST). *International Journal of*
50 *Epidemiology*, 2013, **42**:1620-1628.
51
52
53
54 26. Rodgers SE, Heaven M, Lacey A, *et al*. Cohort profile: the housing regeneration and
55 health study. *Int J Epidemiol* 2012, dys200.
56
57
58
59
60

- 1
2
3 27. SAIL Databank: Core Data Sets. Swansea University: Swansea
4
5 <http://www.saildatabank.com/data-dictionary/sail-datasets> Accessed September 2014.
6
7
8 28. Brock A, Baker A, Griffiths C, *et al.* Suicide trends and geographical variations in the
9
10 United Kingdom, 1991–2004. *Health Statistics Quarterly* 2006, **31**:6-22.
11
12
13 29. Carroll R, Hawton K, Kapur N, *et al.* Impact of the growing use of narrative verdicts
14
15 by coroners on geographic variations in suicide: analysis of coroners' inquest data.
16
17 *Journal of Public Health* 2011, **10**:1-7.
18
19
20 30. Gunnell D, Hawton, K, Kapur N. Coroners' verdicts and suicide statistics in England
21
22 and Wales. *BMJ* 2011, **343**:d6030.
23
24
25 31. Hill C, Cook, L. Narrative verdicts and their impact on mortality statistics in England
26
27 and Wales. *Health Statistics Quarterly* 2011, **49**:81-100.
28
29
30 32. Gunnell D, Bennewith O, Simkin S, *et al.* Time trends in coroners' use of different
31
32 verdicts for possible suicides and their impact on officially reported incidence of
33
34 suicide in England: 1990-2005. *Psychol Med* 2013, **43**:1415-1422.
35
36
37 33. Office of National Statistics. Statistical bulletin: Suicide rates in the United Kingdom,
38
39 2011. ONS 22 January 2013. [http://www.ons.gov.uk/ons/rel/subnational-](http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2011/stb-suicide-bulletin.html)
40
41 [health4/suicides-in-the-united-kingdom/2011/stb-suicide-bulletin.html](http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2011/stb-suicide-bulletin.html) Accessed
42
43 August 2014.
44
45
46 34. Grimes DA, Schultz KF. Compared to what? Finding controls in a case control study.
47
48 *Lancet* 2005, **365**: 1429–33.
49
50
51 35. Office for National Statistics. Welsh index of Multiple Deprivation 2011. Summary
52
53 report, Welsh Government; 2011.
54
55
56
57
58
59
60

- 1
2
3 36. Townsend P, Phillimore P, Beattie A. *Health and Deprivation: Inequality and the*
4
5 *North*. Routledge: London; 1988.
6
- 7 37. Office for National Statistics. Super Output Areas 2011.
8
9 <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super->
10 [output-areas--soas-/index.html](http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-) Accessed August 2014.
11
12
- 13 38. Bruzzi P, Green SB, Byar DP, *et al*. Estimating the population attributable risk for
14 multiple risk factors using case-control data. *Am J Epidemiol* 1985, **122**:904-914.
15
16
- 17 39. Great Britain: Data Protection Act 1998. London: HMSO.
18
19 <http://www.legislation.gov.uk/ukpga/1998/29/contents> Accessed August 2014.
20
21
22
- 23 40. MRC: Policy and Guidance on Sharing of Research Data from Population and Patient
24 Studies. MRC 23rd November 2011.
25
26 <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC008302> Accessed
27 August 2014.
28
29
- 30 41. Wellcome Trust: Data sharing policy. Wellcome Trust August 2010.
31
32 <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position->
33 [statements/WTX035043.htm](http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-) Accessed August 2014.
34
35
36
37
38
39
- 40 42. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic
41 patient record data in primary care. *BMJ* 2003, **326**:1070.
42
43
44
- 45 43. University of Bristol, Suicide Research Centre.
46
47 <http://www.bristol.ac.uk/psychiatry/research/suicideresearch.html> Accessed August
48 2014.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 44. The University of Manchester Centre for Suicide Prevention.

4
5 <http://www.bbmh.manchester.ac.uk/cmhr/research/centreforsuicideprevention/>

6
7 Accessed August 2014.

8
9
10 45. University of Oxford, Centre for Suicide Research. <http://cebmh.warne.ox.ac.uk/csr/>

11
12 Accessed August 2014.

13
14
15 46. ISD Scotland. Scottish Suicide Information Database (ScotSID). ISD Scotland 2011.

16
17 <http://data.gov.uk/dataset/scottish-suicide-information-database> Accessed August

18
19 2014.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —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 <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants	209-282
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	
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) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed	236-282

1
2 *Case-control study*—If applicable, explain how matching of cases
3 and controls was addressed

4 *Cross-sectional study*—If applicable, describe analytical methods
5 taking account of sampling strategy
6

7 (e) Describe any sensitivity analyses

8 Continued on next page
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

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) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	Will
		<i>Cross-sectional study</i> —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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 **Suicide Information Database- Cymru: A protocol for a population-based, routinely**
4 **collected data linkage study to explore risks and patterns of healthcare contact prior to**
5 **suicide to identify opportunities for intervention.**
6
7
8

9
10 John A^{1*}, Dennis M¹, Kosnes L¹, Gunnell D², Scourfield J³, Ford DV¹, Lloyd K¹
11

12
13 ¹College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park,
14 Swansea, SA2 8PP, United Kingdom.
15

16
17
18 ²School of Social and Community Medicine, University of Bristol, Canynge Hall, 39
19 Whatley Road, Bristol, BS8 2PS, United Kingdom.
20

21
22
23 ³School of Social Sciences, Cardiff University, Glamorgan Building, King Edward VII
24 Avenue, Cardiff, CF10 3WT, United Kingdom.
25
26
27

28
29
30
31
32 *Corresponding Author: Ann John^{1*} a.john@swansea.ac.uk; Phone: +44(0)1792 602568;
33
34 College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park,
35 Swansea, SA2 8PP, United Kingdom.
36
37

38
39 Michael Dennis¹ m.s.dennis@swansea.ac.uk, Liv Kosnes¹ l.kosnes@swansea.ac.uk, David
40 Gunnell² d.j.gunnell@bristol.ac.uk, Jonathan Scourfield³ Scourfield@cardiff.ac.uk, David V.
41 Ford¹ d.v.ford@swansea.ac.uk, Keith Lloyd¹ k.r.lloyd@swansea.ac.uk
42
43
44
45
46

47 **Key words:** Suicide, case-control, database, routine data linkage.
48

49
50 **Word Count: 4054**
51
52
53
54
55
56
57

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

1
2
3 stakeholders and national/ international conference presentations. The improved
4
5 understanding of the prior health, nature of previous contacts with services, and wider social
6
7 circumstances of those who complete suicide will assist in prevention policy, service
8
9 organisation and delivery. SID-Cymru is funded through the National Institute for Social
10
11 Care and Health Research, Welsh Government (RFS-12-14).
12

13 14 15 **Article Summary**

16 17 18 Article Focus:

- 19
20 • Case definitions of probable and possible suicides
- 21
22 • Identifying risk factors and patterns of service use in those who complete suicide
- 23
24 • Designing a study
- 25
26
27
28

29 30 Key messages:

- 31
32 • SID-Cymru will facilitate a series of case- control studies of those who complete
33
34 suicide
- 35
36 • Recency of primary care, in- and out-patient hospital and emergency department
37
38 contacts including attendance for self-harm, primary care diagnosis of depression,
39
40 levels of treatment with anti-depressants and trends in such treatment over time, rural
41
42 and urban geography, educational attainment, contacts for the elderly and levels of
43
44 physical illness will be investigated
- 45
46 • Results will be adjusted for a number of confounders
- 47
48
49
50

51 52 Strengths and Limitations of this study:

- 53
54 • Based on the general population
- 55
56
57
58

- 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

For peer review only

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 essential¹. 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 Others¹¹⁻¹³ 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.

1
2
3 In the United Kingdom various systems exist to examine suicide deaths. The National
4 Confidential Inquiry into Suicide and Homicide (NCISH) by people with Mental Illness
5 focuses on suicide cases that were in contact with mental health services (approximately 25%
6 of total) in the year prior to their deaths.³ However this provides limited information on
7 issues of suicide in the general population and may hamper the effectiveness of wider
8 preventive efforts.¹ In a recent report NCISH¹⁴ examined aspects of primary healthcare prior
9 to all suicides in England between 2002-2011; however no linkage was made with data from
10 other service providers such as emergency departments.
11
12
13
14
15
16
17
18
19

20
21 Scotland has recently established enhanced data collection in relation to suicide, however,
22 further development is needed in order for 'ScotSID' to be able to examine healthcare
23 pathways and contact with more than one health service;¹⁵ England does not currently have a
24 dedicated repository for suicide data though studies have made use of the Clinical Practice
25 Research Datalink (CPRD; www.cprd.com), which represents approximately 8.5% of the UK
26 population from 600 General Practices (GP's) in England.¹⁶ CPRD can be linked with data
27 from the National Health Service (NHS) Hospital Episode Statistics (HES;
28 <http://www.hscic.gov.uk>) and mortality data from the Office of National Statistics (ONS;
29 <http://www.ons.gov.uk>) but has limited emergency department data.
30
31
32
33
34
35
36
37
38
39
40
41

42 SID-Cymru will access and link information on prior health, nature of previous contacts with
43 services, and wider social circumstances of all those who complete suicide (known and
44 unknown to mental health services) within the population of Wales via anonymised routinely
45 collected electronic data held in healthcare and social datasets from the Secure Anonymised
46 Information Linkage (SAIL) Databank (www.saildatabank.com).^{17 18} This brings together,
47 links and anonymises the widest possible range of person-based data currently available in
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 the UK. The SAIL Databank was originally set up by the Health Information Research Unit
4
5 (HIRU) at the College of Medicine at Swansea University.
6
7

8
9 SID-Cymru is part of the research programme related to the Health e-Research Collaboration
10
11 UK (HeRC UK), led by the Medical Research Council (MRC) and based in the Centre for the
12
13 Improvement of Population Health through e-Records Research (CIPHER). CIPHER is a UK
14
15 Clinical Research Collaboration (UKCRC) Public Health Research Centre of Excellence set
16
17 within the College of Medicine at Swansea University. SID-Cymru will provide timely robust
18
19 data to inform the future strategic direction of and the first step in designing and evaluating
20
21 effective interventions to prevent suicide. It will use the International Classification of
22
23 Disease, Tenth Revision,¹⁹ (ICD-10) definitions and instructions for classifying causes of
24
25 death which will allow for comparisons with other countries and thus support research of
26
27 relevance globally. An emphasis will be given to issues where there are opportunities for
28
29 intervention or where electronic data linkage confers an advantage to its investigation.
30
31

32 33 **Aim**

34
35
36 The main aim is to establish a Suicide Information Database- Cymru as a population based
37
38 resource for studying factors and service contacts associated with all suicide deaths through
39
40 routinely collected data linkage case control studies.
41
42

43 44 **Objectives:**

45 46 *Phase 1:*

- 47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
 - 50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

- 1
2
3
4
5
6
7
8
9
- 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.

10
11

Phase 2:

- 12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
- 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.

27
28

Methods and Analyses

29
30
31

Design

32
33
34
35
36
37
38
39
40
41
42
43
44

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 million²⁰ 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

45
46
47

Data Source

48
49
50
51
52
53
54
55
56

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

1
2
3 Within the SAIL Databank a split-file approach to anonymisation is used to overcome issues
4 of confidentiality and disclosure in health-related data warehousing.¹⁸ Demographic data are
5 sent to a partner organisation, NHS Wales Information Service (NWIS), where all identifiable
6 information is removed; clinical data are sent directly to HIRU, where, for each dataset
7 within the SAIL Databank, an individual is assigned an encrypted Anonymised Linking Field
8 (ALF). The ALF is used to link anonymised individuals across datasets, thus supporting the
9 opportunity to conduct longitudinal analyses of an individual's journey through multiple
10 health, education and social datasets.¹⁷ Additionally Residential Anonymised Linking Fields
11 (RALFs) have been created for all residences in Wales and enables linkage of anonymised
12 household and environment data with the health records of individual residents without the
13 identity of the residences or residents being known to researchers.²⁶

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28 *The primary study base* will be the Welsh Demographic Service²⁷ (WDS) hosted within the
29 SAIL Databank. The WDS is a core dataset available within the SAIL Databank and part of a
30 set of services to manage administrative information (demographic data) for NHS patients in
31 Wales. The WDS was introduced early in 2009 replacing a similar service known as the NHS
32 Wales Administrative Register (NHS AR). The WDS data is collected from GP practices via
33 the Exeter System and more than five million individuals are currently present in the WDS
34 dataset within the SAIL Databank. The WDS is a register of all individuals that have at some
35 point in time been registered with a Welsh General Practitioner (GP) or required some form
36 of NHS health care provision in Wales. The electronic collation of WDS / NHS AR data
37 originated in 1960 and is updated and maintained by NWIS,²⁷ ensuring that address changes
38 (within and out of Wales) and death notices are included in the register. The original (non-
39 anonymised) version of the NHS AR has been used in the HIRU Matching Algorithm for
40 Consistent Results in Anonymised Linkage (MACRAL), making the WDS / NHS AR the
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 master list for all Welsh residents and using probabilistic matching to find the associated
4
5 NHS numbers that are then encrypted into ALF's.
6
7

8 Deaths in Wales should be registered within five days of the date of death. However
9
10 legislation in Wales means that when a coroner's inquest takes place, the death cannot be
11
12 registered until the inquest is complete. Since the Office for National Statistics (ONS), the
13
14 national agency where all deaths are collated, has no conclusive information about the death
15
16 until it is registered, there is a delay between the date the death occurred and when the death
17
18 is added to the annual ONS mortality dataset. The ONS found that less than 41% of deaths
19
20 going to inquest were registered within 3 months, though 96% were registered within 1
21
22 year.²⁸ Thus precise information on the annual incidence of suicide may be delayed by up to
23
24 two years. Information collected at death registration is recorded on the Registration Online
25
26 system by registrars. Most of the information is normally supplied by the informant (usually a
27
28 close relative of the deceased) while the cause of death (COD) is usually obtained from the
29
30 Medical Certificate of Cause of Death (MCCD) completed by a medical practitioner when
31
32 the death is certified using ICD-10 coding or the coroner if there is an inquest.¹⁹ Notably, a
33
34 death is not officially registered within the Annual District Deaths Extract (ADDE) until the
35
36 COD has been finalised, and thus the year of death and the year of registration may not
37
38 concur. The *primary dataset* used to construct SID-Cymru is the ADDE from the ONS. The
39
40 ADDE is inclusive of Welsh residents who died outside of Wales, and holds information
41
42 about COD derived from death certificates on all deaths in Wales.
43
44
45
46
47
48

49 *Definition of Suicide for Cases*

50
51

52 The true number of suicides is difficult to determine because a coroner's conclusion of
53
54 suicide must be 'beyond a reasonable doubt' i.e. that the death was intentionally self inflicted
55
56 and in some areas coroners have increasingly (since 2001) reported narrative conclusions
57
58

1
2
3 rather than reporting it as suicide.^{29 30} Previously when insufficient information was recorded
4
5 by the coroner ONS coders used to record the death as an accident which inevitably led to
6
7 some suicides being classified as accidents or misadventure. The ONS has recently issued
8
9 guidance on this issue following a coding practice review.³¹ Current ONS practice includes
10
11 deaths where intention is ‘undetermined whether accidentally or purposefully inflicted’; thus
12
13 deaths where there may be no intention to take life, such as in relation to injury or poisoning,
14
15 are included in suicide figures by ONS. Currently there is no access to coroners’ narrative
16
17 verdicts within the SAIL Databank as a possible method for review of case inclusion.
18
19

20
21
22 There is evidence to suggest that a high proportion of deaths from poisoning and hanging that
23
24 receive accidental verdicts, are found, when subjected to clinical review, to be suicides.³²
25
26 Such possible deaths through suicide will be included in SID-Cymru as an opportunity for
27
28 further separate and combined analysis; thus the additional ICD-10 codes relating to
29
30 ‘accidental poisoning with prescribed drugs’ (X40-X41, X43-X49) and ‘accidental hanging’
31
32 (W75-W76) may be used along with ‘sequelae of external causes of morbidity and mortality’
33
34 (Y87, Y87.2, Y89, Y89.9).¹⁸ Thus with SID-Cymru we aim to establish a Suicide
35
36 Information Database that can be a resource for studying factors associated with *all*
37
38 probable, as well as, possible deaths through suicide; by including accidental
39
40 hanging/strangulation and accidental poisoning, excluding, narcotics and psychodysleptics
41
42 (i.e., possible suicides), along with the ONS traditional method of defining suicides, which
43
44 includes suicides and deaths of undetermined intent (probable suicides).²
45
46
47

48 49 *Identification of Cases and Controls*

50
51
52 SID-Cymru cases encompass suicides (‘intentional self-harm’ [ICD-10: X60-84]) and
53
54 probable deaths through suicide (‘undetermined intent’ [ICD-10: Y10-34, excl.Y33.9])
55
56 recorded from the MCCD and presented within ADDE as the underlying COD. Those coded
57
58
59
60

1
2
3 Y33.9/U50.9 (pending verdicts) are excluded since a large proportion of these are
4
5 subsequently found to be homicides. Probable suicide defined above can be supplemented
6
7 with possible deaths through suicide if required. Cases of probable suicide will be identified
8
9 and extracted by use of the ICD-10 codes¹⁹ defined above and depicted in Table 1.
10
11

12
13 ONS figures include those over 15 years of age only, due to the possibility that deaths in
14
15 younger children coded as undetermined events may be caused by unverifiable accidents,
16
17 neglect or abuse.³³ SID-Cymru will allow for the analysis of suicides and probable suicides
18
19 for the 10 to 14 year age band. Official ONS mortality statistics are produced based on the
20
21 number of deaths registered in a particular calendar year, rather than the number of deaths
22
23 that occurred in that year. This means their figures include some deaths that occurred in years
24
25 prior to the reference year (approximately 4%). As SID-Cymru will link and review data in
26
27 the period leading up to death it is important to ensure the match/end date reflects the correct
28
29 time period (specific to each individual case), i.e. date of death not date of registration, to
30
31 afford an accurate perspective on utilisation of resources and help seeking behaviours.
32
33 Consequently the actual date of death (ADOD) will be used in the matching criteria to
34
35 establish a data review 'end date' for controls, rather than the registered date of death
36
37 (RDOD) referred to in ONS reports. Mortality data within the SAIL Databank is only
38
39 available from 2003, thus the earliest case inclusion relates to ADOD's from the 1st January
40
41 2003. To minimise the underestimation of cases identified (i.e., due to delays in COD
42
43 confirmation and registration of DOD) the case inclusion for SID-Cymru includes cases aged
44
45 10 years and over at ADOD, where the ADOD took place between 01/01/2003 and 31/12/
46
47 2011.
48
49
50
51
52
53
54
55
56
57
58
59
60

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 Suicide¹ 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 ADOD³ within SID⁴	361	303	287	278	279	282	272	292	304	2664
Probable Suicides RDOD⁵ within SID	274	330	270	300	288	266	257	287	340	2612
ONS Probable Suicides⁶ RDOD	332	330	269	300	289	266	258	288	341	2673

¹Probable suicides = ICD-10 codes for Intentional [X60-84], Undetermined [Y10-34 excl. Y33.9] self-harm;

²Values are number of individuals unless otherwise indicated ³ADOD = Year death occurred; ⁴Cases identified for SID-Cymru includes those aged 10 years and over; ⁵RDOD = Year death was registered; ⁶Cases 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 Deprivation³⁵ (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

1
2
3 A case-control study utilising SID-Cymru will be population based and so the relative risk of
4
5 suicide will be estimated by conditional logistic regression model with SPSS (version 20).
6
7 Crude odds ratios will be adjusted for general practice and/or LSOA by matching cases and
8
9 controls. Unadjusted estimates, confounder-adjusted estimates and their precision (e.g., 95%
10
11 confidence interval) will be produced. Interactions between variables will be assessed with
12
13 the log likelihood ratio test based on results from the adjusted analysis. The population
14
15 attributable risk will be calculated³⁸ on the basis of adjusted relative risks from the full
16
17 analysis and the distribution of exposures in the cases.
18
19

20
21 We will also report information on the completeness of linkage with each dataset.
22
23

24 **Ethics and Dissemination**

25 *Ethics*

26
27
28 A large amount of preliminary work on anonymisation methodologies was undertaken to
29
30 create the SAIL Databank system,^{17- 18} and the SAIL Databank has the required ethical
31
32 permissions and processes in place to analyse anonymised data. It operates within a robust
33
34 series of guidelines in line with the Caldicott principles and the National Information
35
36 Governance Board for Health and Social Care.¹⁸ Ethical approval has been granted for SID-
37
38 Cymru from HIRU's Information Governance Review Panel (IGRP), an independent body
39
40 consisting of a range of government, regulatory and professional agencies. In compliance
41
42 with the IGRP rulings and the Data Protection Act 1998³⁹ individual level data and personal
43
44 identifier linkage codes will not be removed from the SAIL Databank and all analyses will be
45
46 carried out within the SAIL Databank gateway at Swansea University, a secure access point
47
48 to data within the SAIL Databank. The key points of the MRC/Wellcome Trust data sharing
49
50 policy will be followed.^{40 41}
51
52
53
54
55
56
57
58
59
60

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

1
2
3 aid colleagues who may lack such analytical expertise and foster greater multi-disciplinary
4
5 collaborations and advance suicide research.
6
7

8 The UK has a strong presence in form of a wide range of publications and expertise relating
9
10 to suicide research. Successful and dedicated Suicide Research Centres exist in Bristol,
11
12 Manchester and Oxford,⁴³⁻⁴⁵ and Scotland recently commenced work on a Scottish Suicide
13
14 Information Database ('ScotSID').⁴⁶ These centres of excellence report broadly about suicide,
15
16 though for logistical reasons, often a regional focus is retained, relating to project funding,
17
18 analytical and interventional work; for that reason the establishment of a Suicide Information
19
20 Database in Wales will build on and enhance existing UK suicide resources and infrastructure
21
22 since the level of linkage available within the SAIL Databank is unique in this field. Unique
23
24 opportunities within the SAIL Databank include linkage across primary, secondary and
25
26 emergency department data and with education data. Findings will be disseminated through
27
28 publications in peer reviewed journals and presentations at local, national and international
29
30 conferences. Communication and consultation with key stakeholders from health and social
31
32 care (for example primary care, mental health, Royal Colleges), government and other policy
33
34 makers, as well as, the third sector will occur. Dissemination will be facilitated by the wider
35
36 roles and responsibilities in suicide prevention, nationally and internationally, of members of
37
38 the research team.
39
40
41
42
43

44 **Implications and Significance**

45
46 Valuable opportunities exist for a wide range of epidemiological and clinical studies on
47
48 suicide in Wales and SID-Cymru has the potential to become an important resource in
49
50 facilitating such research which will be of relevance internationally. In addition to the records
51
52 that have already been included for linkage with SID-Cymru it is expected that over time
53
54 relevant information from other data sources (e.g. the DWP) will be linked to SID-Cymru to
55
56
57
58

1
2
3 provide a wider range of information on issues such as individuals' circumstances, the nature
4 of their deaths, and their contact with extended services. Additionally, linkage can be made
5 with non-routinely collected datasets such as that held by NCISH. Specific hypotheses that
6 will be explored include: recency of primary care, hospital and emergency department
7 contacts including attendance for self-harm, primary care diagnosis of depression, levels of
8 treatment with anti-depressants and trends in such treatment over time, rural and urban
9 geography, contacts for the elderly and levels of physical illness.

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

22 **Authors' contributions**

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
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.

1
2
3 [http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/Ann](http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/AnnualReport2013_UK.pdf)
4
5 [ualReport2013_UK.pdf](http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/AnnualReport2013_UK.pdf) Accessed July 2014.
6
7

- 8 4. Andersen UA, Andersen M, Rosholm JU, *et al.* Contacts to the health care system
9 prior to suicide: a comprehensive analysis using registers for general and psychiatric
10 hospital admissions, contacts to general practitioners and practising specialists and
11 drug prescriptions. *Acta Psychiatr Scand* 2000, **102**:126-34.
12
- 13 5. Qin P, Agerbo E, Mortensen PB. Suicide risk in relation to socioeconomic,
14 demographic, psychiatric, and familial factors: a national register-based study of all
15 suicides in Denmark, 1981–1997. *Am J Psychiatry* 2003, **160**:765-772.
16
- 17 6. Gunnell D, Löfving S, Gustafsson JE, *et al.* School performance and risk of suicide in
18 early adulthood: follow-up of two national cohorts of Swedish schoolchildren.
19
20 *Journal of affective disorders*, 2011, **131**:104-112.
21
- 22 7. Qin P, Hawton K, Mortensen PB, *et al.* Combined effects of physical illness and
23 comorbid psychiatric disorder on risk of suicide in a national population study. *Br J*
24 *Psychiatry* 2014, **204**:430-435.
25
- 26 8. Christiansen E, Larsen KJ, Agerbo E, *et al.* Incidence and risk factors for suicide
27 attempts in a general population of young people: a Danish register-based study. *Aust*
28 *N Zealand J Psychiatry* 2012, **47**: 259-270.
29
- 30 9. Björkenstam C, Johansson LA, Nordström P, *et al.* Suicide or undetermined intent? A
31 register-based study of signs of misclassification. *Population health metrics*, 2014,
32
33 **12**:11.
34
- 35 10. Hjorthøj CR, Madsen T, Agerbo E, *et al.* Risk of suicide according to level of
36 psychiatric treatment: a nationwide nested case-control study. *Social psychiatry and*
37 *psychiatric epidemiology*, 2014; 1-9.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 11. Lawrence DM, Holman CD, Jablensky AV, *et al.* Suicide rates in psychiatric in-
4
5 patients: an application of record linkage to mental health research. *Aust NZ J Public*
6
7 *Health* 1999, **23**:468–470.
- 8
9
10 12. Lawrence D, D'Arcy C, Holman J, *et al.* Increasing rates of suicide in Western
11
12 Australian psychiatric patients: a record linkage study. *Acta Psychiatr Scand* 2001,
13
14 **104**:443-451.
- 15
16
17 13. Morrison KB, Laing L. *Adults' use of health services in the year before death by*
18
19 *suicide in Alberta.* Statistics Canada; 2011.
- 20
21 14. Appleby L, Kapur N, Shaw J, *et al.* Suicide in primary care in England: 2002-2011.
22
23 National Confidential Inquiry into Suicide and Homicide by People with Mental
24
25 Illness (NCISH). Manchester: University of Manchester 2014.
26
27 <http://www.bbmh.manchester.ac.uk/cmhr/research/centreforsuicideprevention/nci/reports/SuicideinPrimaryCare2014.pdf> Accessed August 2014.
- 28
29
30
31
32 15. ISD Scotland: The Scottish Suicide Information Database Report 2014.
33
34 <http://www.isdscotland.org/Health-Topics/Public-Health/Publications/index.asp?ID=1232> Accessed August 2014.
- 35
36
37
38
39 16. Webb RT, Kontopantelis E, Doran T, *et al.* Suicide risk in primary care patients with
40
41 major physical diseases: a case-control study. *Arch Gen Psychiatry* 2012, **69**:256-264.
- 42
43
44 17. Lyons RA, Jones KH, John G, *et al.* The SAIL databank: linking multiple health and
45
46 social care datasets. *BMC Med Inform Decis Making* 2009, **9**:3.
- 47
48
49
50 18. Ford DV, Jones KH, Verplancke JP, *et al.* The SAIL Databank: building a national
51
52 architecture for e-health research and evaluation. *BMC Health Serv Res* 2009, **9**:1.
- 53
54
55
56
57
58
59
60

- 1
2
3 19. World Health Organization: *The ICD-10 classification of mental and behavioural*
4 *disorders: Clinical descriptions and diagnostic guidelines*. WHO: Geneva; 1993.
5
6
7
8 20. Office of National Statistics. 2011 Census Results for Wales. ONS 16th July 2012.
9
10 <http://wales.gov.uk/docs/statistics/2012/120716-census-2011-first-results-wales->
11 [en.pdf](http://wales.gov.uk/docs/statistics/2012/120716-census-2011-first-results-wales-en.pdf) Accessed September 2014.
12
13
14
15 21. Office of National Statistics. Statistical bulletin: Suicide rates in the United Kingdom,
16 2006 to 2010. ONS 26th January 2012. [http://www.ons.gov.uk/ons/rel/subnational-](http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2010/stb-statistical-bulletin.html)
17 [health4/suicides-in-the-united-kingdom/2010/stb-statistical-bulletin.html](http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2010/stb-statistical-bulletin.html) Accessed
18 August 2014.
19
20
21
22
23
24
25 22. McGregor JJ, Brooks CJ, Chalasani P, *et al*. The Health Informatics Trial
26 Enhancement Project (HITE): Using routinely collected primary care data to identify
27 potential participants for a depression trial. *Trials* 2010, **11**:39.
28
29
30
31
32
33 23. Brooks CJ, Stephens JW, Price DE, *et al*. Use of a patient linked data warehouse to
34 facilitate diabetes trial recruitment from primary care. *Primary Care Diabetes* 2009,
35 **3**:245-248
36
37
38
39
40 24. Hill RA, Brophy S, Brunt H, *et al*. Protocol of the baseline assessment for the
41 Environments for Healthy Living (EHL) Wales cohort study. *BMC Public Health*
42 2010, **10**:150
43
44
45
46
47
48 25. Fone DL, Dunstan F, White J, *et al*. Cohort Profile: The Caerphilly Health and Social
49 Needs Electronic Cohort Study (E-CATALyST). *International Journal of*
50 *Epidemiology*, 2013, **42**:1620-1628.
51
52
53
54
55 26. Rodgers SE, Heaven M, Lacey A, *et al*. Cohort profile: the housing regeneration and
56 health study. *Int J Epidemiol* 2012, dys200.
57
58
59
60

- 1
2
3 27. SAIL Databank: Core Data Sets. Swansea University: Swansea
4
5 <http://www.saildatabank.com/data-dictionary/sail-datasets> Accessed September 2014.
6
7
8 28. Brock A, Baker A, Griffiths C, *et al.* Suicide trends and geographical variations in the
9
10 United Kingdom, 1991–2004. *Health Statistics Quarterly* 2006, **31**:6-22.
11
12
13 29. Carroll R, Hawton K, Kapur N, *et al.* Impact of the growing use of narrative verdicts
14
15 by coroners on geographic variations in suicide: analysis of coroners' inquest data.
16
17 *Journal of Public Health* 2011, **10**:1-7.
18
19
20
21 30. Gunnell D, Hawton, K, Kapur N. Coroners' verdicts and suicide statistics in England
22
23 and Wales. *BMJ* 2011, **343**:d6030.
24
25
26 31. Hill C, Cook, L. Narrative verdicts and their impact on mortality statistics in England
27
28 and Wales. *Health Statistics Quarterly* 2011, **49**:81-100.
29
30
31 32. Gunnell D, Bennewith O, Simkin S, *et al.* Time trends in coroners' use of different
32
33 verdicts for possible suicides and their impact on officially reported incidence of
34
35 suicide in England: 1990-2005. *Psychol Med* 2013, **43**:1415-1422.
36
37
38 33. Office of National Statistics. Statistical bulletin: Suicide rates in the United Kingdom,
39
40 2011. ONS 22 January 2013. [http://www.ons.gov.uk/ons/rel/subnational-](http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2011/stb-suicide-bulletin.html)
41
42 [health4/suicides-in-the-united-kingdom/2011/stb-suicide-bulletin.html](http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2011/stb-suicide-bulletin.html) Accessed
43
44 August 2014.
45
46 34. Grimes DA, Schultz KF. Compared to what? Finding controls in a case control study.
47
48 *Lancet* 2005, **365**: 1429–33.
49
50
51 35. Office for National Statistics. Welsh index of Multiple Deprivation 2011. Summary
52
53 report, Welsh Government; 2011.
54
55
56
57
58
59
60

- 1
2
3 36. Townsend P, Phillimore P, Beattie A. *Health and Deprivation: Inequality and the*
4
5 *North*. Routledge: London; 1988.
6
- 7 37. Office for National Statistics. Super Output Areas 2011.
8
9 <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super->
10 [output-areas--soas-/index.html](http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-) Accessed August 2014.
11
12
- 13 38. Bruzzi P, Green SB, Byar DP, *et al*. Estimating the population attributable risk for
14
15 multiple risk factors using case-control data. *Am J Epidemiol* 1985, **122**:904-914.
16
17
- 18 39. Great Britain: Data Protection Act 1998. London: HMSO.
19
20 <http://www.legislation.gov.uk/ukpga/1998/29/contents> Accessed August 2014.
21
22
- 23 40. MRC: Policy and Guidance on Sharing of Research Data from Population and Patient
24
25 Studies. MRC 23rd November 2011.
26
27 <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC008302> Accessed
28
29 August 2014.
30
31
- 32 41. Wellcome Trust: Data sharing policy. Wellcome Trust August 2010.
33
34 <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position->
35 [statements/WTX035043.htm](http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-) Accessed August 2014.
36
37
38
- 39 42. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic
40
41 patient record data in primary care. *BMJ* 2003, **326**:1070.
42
43
44
- 45 43. University of Bristol, Suicide Research Centre.
46
47 <http://www.bristol.ac.uk/psychiatry/research/suicideresearch.html> Accessed August
48
49 2014.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 44. The University of Manchester Centre for Suicide Prevention.

4
5 <http://www.bbmh.manchester.ac.uk/cmhr/research/centreforsuicideprevention/>

6
7 Accessed August 2014.

8
9
10 45. University of Oxford, Centre for Suicide Research. <http://cebmh.warne.ox.ac.uk/csr/>

11
12 Accessed August 2014.

13
14
15 46. ISD Scotland. Scottish Suicide Information Database (ScotSID). ISD Scotland 2011.

16
17 <http://data.gov.uk/dataset/scottish-suicide-information-database> Accessed August

18
19 2014.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7 **Suicide Information Database- Cymru: A protocol for a population-based, routinely**
8 **collected data linkage ~~case-control~~ study to explore risks and patterns of healthcare**
9 **contact prior to suicide to identify opportunities for intervention.**
10
11

12
13 John A^{1*}, Dennis M¹, Kosnes L¹, Gunnell D², Scourfield J³, Ford DV¹, Lloyd K¹

14
15
16 ¹College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park,
17 Swansea, SA2 8PP, United Kingdom.

18
19
20 ²School of Social and Community Medicine, University of Bristol, Canynge Hall, 39
21 Whatley Road, Bristol, BS8 2PS, United Kingdom.

22
23
24 ³School of Social Sciences, Cardiff University, Glamorgan Building, King Edward VII
25 Avenue, Cardiff, CF10 3WT, United Kingdom.
26
27
28

29
30
31
32 *Corresponding Author: Ann John^{1*} a.john@swansea.ac.uk; Phone: +44(0)1792 602568;
33
34 College of Medicine, Swansea University, Institute of Life Sciences 2, Singleton Park,
35 Swansea, SA2 8PP, United Kingdom.
36

37
38 Michael Dennis¹ m.s.dennis@swansea.ac.uk, Liv Kosnes¹ l.kosnes@swansea.ac.uk, David
39 Gunnell² d.j.gunnell@bristol.ac.uk, Jonathan Scourfield³ Scourfield@cardiff.ac.uk, David V.
40 Ford¹ d.v.ford@swansea.ac.uk, Keith Lloyd¹ k.r.lloyd@swansea.ac.uk
41
42
43
44

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

46
47 **Word Count: 4054**
48
49
50
51
52
53
54

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).

Page 2 of 30

1
2
3
4
5
6
7 Findings will be disseminated through peer reviewed publications, consultations with
8 stakeholders and national/ international conference presentations. The improved
9
10 understanding of the prior health, nature of previous contacts with services, and wider social
11
12 circumstances of those who complete suicide will assist in prevention policy, service
13
14 organisation and delivery. SID-Cymru is funded through the National Institute for Social
15
16 Care and Health Research, Welsh Government (RFS-12-14).
17

Formatted: Font: Times New Roman, 12 pt

18 19 **Article Summary**

20 21 Article Focus:

- 22
23 • Case definitions of probable and possible suicides
- 24
25 • Identifying risk factors and patterns of service use in those who complete suicide
- 26
27 • Designing a study
- 28
29
- 30

31 32 Key messages:

- 33
34 • SID-Cymru will facilitate a series of case- control studies of those who complete
35
36 suicide
- 37
38 • Recency of primary care, in- and out-patient hospital and emergency department
39
40 contacts including attendance for self-harm, primary care diagnosis of depression,
41
42 levels of treatment with anti-depressants and trends in such treatment over time, rural
43
44 and urban geography, educational attainment, contacts for the elderly and levels of
45
46 physical illness will be investigated
- 47
48 • Results will be adjusted for a number of confounders
- 49
50

51 52 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 essential¹. 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 Others¹¹⁻¹³ 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.

1
2
3
4
5
6
7 In the United Kingdom various systems exist to examine suicide deaths. The National
8 Confidential Inquiry into Suicide and Homicide (NCISH) by people with Mental Illness
9 focuses on suicide cases that were in contact with mental health services (approximately 25%
10 of total) in the year prior to their deaths.³ However this provides limited information on
11 issues of suicide in the general population and may hamper the effectiveness of wider
12 preventive efforts.¹ In a recent report NCISH¹⁴ examined aspects of primary healthcare prior
13 to all suicides in England between 2002-2011; however no linkage was made with data from
14 other service providers such as emergency departments.
15
16
17
18
19
20
21

22 Scotland has recently established enhanced data collection in relation to suicide, however,
23 further development is needed in order for 'ScotSID' to be able to examine healthcare
24 pathways and contact with more than one health service;¹⁵ England does not currently have a
25 dedicated repository for suicide data though studies have made use of the Clinical Practice
26 Research Datalink (CPRD; www.cprd.com), which represents approximately 8.5% of the UK
27 population from 600 General Practices (GP's) in England.¹⁶ CPRD can be linked with data
28 from the National Health Service (NHS) Hospital Episode Statistics (HES;
29 <http://www.hscic.gov.uk>) and mortality data from the Office of National Statistics (ONS;
30 <http://www.ons.gov.uk>) but has limited emergency department data.
31
32
33
34
35
36
37
38
39

40 SID-Cymru will access and link information on prior health, nature of previous contacts with
41 services, and wider social circumstances of all those who complete suicide (known and
42 unknown to mental health services) within the population of Wales via anonymised routinely
43 collected electronic data held in healthcare and social datasets from the Secure Anonymised
44 Information Linkage (SAIL) Databank (www.saildatabank.com).^{17 18} This brings together,
45 links and anonymises the widest possible range of person-based data currently available in
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7 the UK. The SAIL Databank was originally set up by the Health Information Research Unit
8 (HIRU) at the College of Medicine at Swansea University.
9

10
11 SID-Cymru is part of the research programme related to the Health e-Research Collaboration
12 UK (HeRC UK), led by the Medical Research Council (MRC) and based in the Centre for the
13 Improvement of Population Health through e-Records Research (CIPHER). CIPHER is a UK
14 Clinical Research Collaboration (UKCRC) Public Health Research Centre of Excellence set
15 within the College of Medicine at Swansea University. SID-Cymru will provide timely robust
16 data to inform the future strategic direction of and the first step in designing and evaluating
17 effective interventions to prevent suicide. It will use the International Classification of
18 Disease, Tenth Revision,¹⁹ (ICD-10) definitions and instructions for classifying causes of
19 death which will allow for comparisons with other countries and thus support research of
20 relevance globally. An emphasis will be given to issues where there are opportunities for
21 intervention or where electronic data linkage confers an advantage to its investigation.
22
23
24
25
26
27
28
29
30
31

32 33 **Aim**

34
35 The main aim is to establish a Suicide Information Database- Cymru as a population based
36 resource for studying factors and service contacts associated with all suicide deaths through
37 routinely collected data linkage case control studies.
38
39
40

41 42 **Objectives:**

43 44 *Phase 1:*

- 45
46
47 ▪ To identify, via the SAIL Databank, all those with ICD-10 codes for probable and
48 possible suicide in Wales 2003-2011, and matched controls.
49
50
51
52
53
54

- 1
2
3
4
5
6
7
8
9
10
11
12
- 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.

13
14

Phase 2:

- 15
16
17
18
19
20
21
22
23
24
25
26
27
- 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.

28
29

Methods and Analyses

30
31
32

Design

33
34
35
36
37
38
39
40
41
42

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 million²⁰ 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

43
44
45

Data Source

46
47
48
49
50
51
52
53
54

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

1
2
3
4
5
6
7 Within the SAIL Databank a split-file approach to anonymisation is used to overcome issues
8 of confidentiality and disclosure in health-related data warehousing.¹⁸ Demographic data are
9 sent to a partner organisation, NHS Wales Information Service (NWIS), where all identifiable
10 information is removed; clinical data are sent directly to HIRU, where, for each dataset
11 within the SAIL Databank, an individual is assigned an encrypted Anonymised Linking Field
12 (ALF). The ALF is used to link anonymised individuals across datasets, thus supporting the
13 opportunity to conduct longitudinal analyses of an individual's journey through multiple
14 health, education and social datasets.¹⁷ Additionally Residential Anonymised Linking Fields
15 (RALFs) have been created for all residences in Wales and enables linkage of anonymised
16 household and environment data with the health records of individual residents without the
17 identity of the residences or residents being known to researchers.²⁶

18
19
20
21
22
23
24
25
26
27
28
29 *The primary study base* will be the Welsh Demographic Service²⁷ (WDS) hosted within the
30 SAIL Databank. The WDS is a core dataset available within the SAIL Databank and part of a
31 set of services to manage administrative information (demographic data) for NHS patients in
32 Wales. The WDS was introduced early in 2009 replacing a similar service known as the NHS
33 Wales Administrative Register (NHS AR). The WDS data is collected from GP practices via
34 the Exeter System and more than five million individuals are currently present in the WDS
35 dataset within the SAIL Databank. The WDS is a register of all individuals that have at some
36 point in time been registered with a Welsh General Practitioner (GP) or required some form
37 of NHS health care provision in Wales. The electronic collation of WDS / NHS AR data
38 originated in 1960 and is updated and maintained by NWIS,²⁷ ensuring that address changes
39 (within and out of Wales) and death notices are included in the register. The original (non-
40 anonymised) version of the NHS AR has been used in the HIRU Matching Algorithm for
41 Consistent Results in Anonymised Linkage (MACRAL), making the WDS / NHS AR the
42
43
44
45
46
47
48
49
50
51
52
53

1
2
3
4
5
6
7 master list for all Welsh residents and using probabilistic matching to find the associated
8 NHS numbers that are then encrypted into ALF's.
9

10
11 Deaths in Wales should be registered within five days of the date of death. However
12 legislation in Wales means that when a coroner's inquest takes place, the death cannot be
13 registered until the inquest is complete. Since the Office for National Statistics (ONS), the
14 national agency where all deaths are collated, has no conclusive information about the death
15 until it is registered, there is a delay between the date the death occurred and when the death
16 is added to the annual ONS mortality dataset. The ONS found that less than 41% of deaths
17 going to inquest were registered within 3 months, though 96% were registered within 1
18 year.²⁸ Thus precise information on the annual incidence of suicide may be delayed by up to
19 two years. Information collected at death registration is recorded on the Registration Online
20 system by registrars. Most of the information is normally supplied by the informant (usually a
21 close relative of the deceased) while the cause of death (COD) is usually obtained from the
22 Medical Certificate of Cause of Death (MCCD) completed by a medical practitioner when
23 the death is certified using ICD-10 coding or the coroner if there is an inquest.¹⁹ Notably, a
24 death is not officially registered within the Annual District Deaths Extract (ADDE) until the
25 COD has been finalised, and thus the year of death and the year of registration may not
26 concur. The *primary dataset* used to construct SID-Cymru is the ADDE from the ONS. The
27 ADDE is inclusive of Welsh residents who died outside of Wales, and holds information
28 about COD derived from death certificates on all deaths in Wales.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

46 *Definition of Suicide for Cases*

47

48
49 The true number of suicides is difficult to determine because a coroner's conclusion of
50 suicide must be 'beyond a reasonable doubt' i.e. that the death was intentionally self inflicted
51 and in some areas coroners have increasingly (since 2001) reported narrative conclusions
52
53
54

1
2
3
4
5
6 rather than reporting it as suicide.^{29 30} Previously when insufficient information was recorded
7
8 by the coroner ONS coders used to record the death as an accident which inevitably led to
9
10 some suicides being classified as accidents or misadventure. The ONS has recently issued
11
12 guidance on this issue following a coding practice review.³¹ Current ONS practice includes
13
14 deaths where intention is ‘undetermined whether accidentally or purposefully inflicted’; thus
15
16 deaths where there may be no intention to take life, such as in relation to injury or poisoning,
17
18 are included in suicide figures by ONS. Currently there is no access to coroners’ narrative
19
20 verdicts within the SAIL Databank as a possible method for review of case inclusion.
21

22
23 There is evidence to suggest that a high proportion of deaths from poisoning and hanging that
24
25 receive accidental verdicts, are found, when subjected to clinical review, to be suicides.³²
26
27 Such possible deaths through suicide will be included in SID-Cymru as an opportunity for
28
29 further separate and combined analysis; thus the additional ICD-10 codes relating to
30
31 ‘accidental poisoning with prescribed drugs’ (X40-X41, X43-X49) and ‘accidental hanging’
32
33 (W75-W76) may be used along with ‘sequelae of external causes of morbidity and mortality’
34
35 (Y87, Y87.2, Y89, Y89.9).¹⁸ Thus with SID-Cymru we aim to establish a Suicide
36
37 Information Database that can be a resource for studying factors associated with *all*
38
39 probable, as well as, possible deaths through suicide; by including accidental
40
41 hanging/strangulation and accidental poisoning, excluding, narcotics and psychodysleptics
42
43 (i.e., possible suicides), along with the ONS traditional method of defining suicides, which
44
45 includes suicides and deaths of undetermined intent (probable suicides).²
46

47 *Identification of Cases and Controls*

48
49 SID-Cymru cases encompass suicides (‘intentional self-harm’ [ICD-10: X60-84]) and
50
51 probable deaths through suicide (‘undetermined intent’ [ICD-10: Y10-34, excl.Y33.9])
52
53 recorded from the MCCD and presented within ADDE as the underlying COD. Those coded
54
55
56
57
58
59
60

1
2
3
4
5
6
7 Y33.9/U50.9 (pending verdicts) are excluded since a large proportion of these are
8 subsequently found to be homicides. Probable suicide defined above can be supplemented
9 with possible deaths through suicide if required. Cases of probable suicide will be identified
10 and extracted by use of the ICD-10 codes¹⁹ defined above and depicted in Table 1.
11
12
13

14
15
16 ONS figures include those over 15 years of age only, due to the possibility that deaths in
17 younger children coded as undetermined events may be caused by unverifiable accidents,
18 neglect or abuse.³³ SID-Cymru will allow for the analysis of suicides and probable suicides
19 for the 10 to 14 year age band. Official ONS mortality statistics are produced based on the
20 number of deaths registered in a particular calendar year, rather than the number of deaths
21 that occurred in that year. This means their figures include some deaths that occurred in years
22 prior to the reference year (approximately 4%). As SID-Cymru will link and review data in
23 the period leading up to death it is important to ensure the match/end date reflects the correct
24 time period (specific to each individual case), i.e. date of death not date of registration, to
25 afford an accurate perspective on utilisation of resources and help seeking behaviours.
26 Consequently the actual date of death (ADOD) will be used in the matching criteria to
27 establish a data review 'end date' for controls, rather than the registered date of death
28 (RDOD) referred to in ONS reports. Mortality data within the SAIL Databank is only
29 available from 2003, thus the earliest case inclusion relates to ADOD's from the 1st January
30 2003. To minimise the underestimation of cases identified (i.e., due to delays in COD
31 confirmation and registration of DOD) the case inclusion for SID-Cymru includes cases aged
32 10 years and over at ADOD, where the ADOD took place between 01/01/2003 and 31/12/
33 2011.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

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 Suicide¹ 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 ADOD³ within SID⁴	361	303	287	278	279	282	272	292	304	2664
Probable Suicides RDOD⁵ within SID	274	330	270	300	288	266	257	287	340	2612
ONS Probable Suicides⁶ RDOD	332	330	269	300	289	266	258	288	341	2673

¹Probable suicides = ICD-10 codes for Intentional [X60-84], Undetermined [Y10-34 excl.Y33.9] self-harm;

²Values are number of individuals unless otherwise indicated ³ADOD = Year death occurred; ⁴Cases identified for SID-Cymru includes those aged 10 years and over; ⁵RDOD = Year death was registered; ⁶Cases 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

1
2
3
4
5
6
7 *Routine Data Sources*
8

9 For SID-Cymru the data collected on identified cases and controls, via ADDE and the WDS
10 respectively, will be linked to other routinely collected datasets, retrospectively allowing a
11 review of each individual's pathway through the various services. Linkage with, for example,
12 GP system data provides varying information about patients going back several years,
13 including previous diagnosis, presenting symptoms and previous medications prescribed.
14 This dataset can be used to review contacts with the GP and, consequently, infer the
15 development / diagnosis of any new medical conditions including depression and self harm
16 prior to suicide. Linkage with in-patient data will allow a review of hospital contacts and
17 Emergency Department datasets will give information on crisis contacts. This will provide
18 comprehensive insights into help seeking behaviours and management across settings. Data
19 sets currently accessible via the SAIL Databank which will be linked to SID-Cymru are
20 presented in Table 3. There are several other datasets currently under negotiation for
21 inclusion within the SAIL Databank to which, in the future, SID-Cymru could link. These
22 include Department of Work and Pensions (DWP) employment and incapacity status, Looked
23 After Children, fostering, Substance Misuse Services, Sexually Transmitted Infections and
24 Police Data.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 Deprivation³⁵ (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

1
2
3
4
5
6
7 A case-control study utilising SID-Cymru will be population based and so the relative risk of
8 suicide will be estimated by conditional logistic regression model with SPSS (version 20).
9
10 Crude odds ratios will be adjusted for general practice and/or LSOA by matching cases and
11 controls. Unadjusted estimates, confounder-adjusted estimates and their precision (e.g., 95%
12 confidence interval) will be produced. Interactions between variables will be assessed with
13
14 the log likelihood ratio test based on results from the adjusted analysis. The population
15
16 attributable risk will be calculated³⁸ on the basis of adjusted relative risks from the full
17
18 analysis and the distribution of exposures in the cases.
19
20

21
22 We will also report information on the completeness of linkage with each dataset.
23
24

25 **Ethics and Dissemination**

26 *Ethics*

27
28
29 A large amount of preliminary work on anonymisation methodologies was undertaken to
30
31 create the SAIL Databank system,^{17- 18} and the SAIL Databank has the required ethical
32
33 permissions and processes in place to analyse anonymised data. It operates within a robust
34
35 series of guidelines in line with the Caldicott principles and the National Information
36
37 Governance Board for Health and Social Care.¹⁸ Ethical approval has been granted for SID-
38
39 Cymru from HIRU's Information Governance Review Panel (IGRP), an independent body
40
41 consisting of a range of government, regulatory and professional agencies. In compliance
42
43 with the IGRP rulings and the Data Protection Act 1998³⁹ individual level data and personal
44
45 identifier linkage codes will not be removed from the SAIL Databank and all analyses will be
46
47 carried out within the SAIL Databank gateway at Swansea University, a secure access point
48
49 to data within the SAIL Databank. The key points of the MRC/Wellcome Trust data sharing
50
51 policy will be followed.^{40 41}
52
53

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

1
2
3
4
5
6
7 aid colleagues who may lack such analytical expertise and foster greater multi-disciplinary
8 collaborations and advance suicide research.
9

10
11 The UK has a strong presence in form of a wide range of publications and expertise relating
12 to suicide research. Successful and dedicated Suicide Research Centres exist in Bristol,
13 Manchester and Oxford,⁴³⁻⁴⁵ and Scotland recently commenced work on a Scottish Suicide
14 Information Database ('ScotSID').⁴⁶ These centres of excellence report broadly about suicide,
15 though for logistical reasons, often a regional focus is retained, relating to project funding,
16 analytical and interventional work; for that reason the establishment of a Suicide Information
17 Database in Wales will build on and enhance existing UK suicide resources and infrastructure
18 since the level of linkage available within the SAIL Databank is unique in this field. Unique
19 opportunities within the SAIL Databank include linkage across primary, secondary and
20 emergency department data and with education data. Findings will be disseminated through
21 publications in peer reviewed journals and presentations at local, national and international
22 conferences. Communication and consultation with key stakeholders from health and social
23 care (for example primary care, mental health, Royal Colleges), government and other policy
24 makers, as well as, the third sector will occur. Dissemination will be facilitated by the wider
25 roles and responsibilities in suicide prevention, nationally and internationally, of members of
26 the research team.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

Formatted: Font: Times New Roman, 12 pt

43 **Implications and Significance**

44
45 Valuable opportunities exist for a wide range of epidemiological and clinical studies on
46 suicide in Wales and SID-Cymru has the potential to become an important resource in
47 facilitating such research which will be of relevance internationally. In addition to the records
48 that have already been included for linkage with SID-Cymru it is expected that over time
49 relevant information from other data sources (e.g. the DWP) will be linked to SID-Cymru to
50
51
52
53
54

1
2
3
4
5
6
7 provide a wider range of information on issues such as individuals' circumstances, the nature
8 of their deaths, and their contact with extended services. Additionally, linkage can be made
9 with non-routinely collected datasets such as that held by NCISH. Specific hypotheses that
10 will be explored include: recency of primary care, hospital and emergency department
11 contacts including attendance for self-harm, primary care diagnosis of depression, levels of
12 treatment with anti-depressants and trends in such treatment over time, rural and urban
13 geography, contacts for the elderly and levels of physical illness.

14
15
16 Findings of current and projected public health importance will be assessed and presented to
17 support policy makers, commissioners and providers of health and social care in Wales. Non-
18 identifiable information from this project will be made available to researchers in Wales, the
19 UK and international collaborators. The initial project focus will be on identification of cases
20 and controls, data linkage opportunities and methodological issues relating to the
21 establishment of a Suicide Information Database and routine data linkage (Phase 1), before
22 commencing data extraction and analysis (Phase 2). The proposed data collation and linkage
23 of primary, secondary and emergency department health information together with
24 educational data are currently unique in the United Kingdom. ~~ThusDue to existing routine~~
25 ~~data linkages within the SAIL Databank~~¹⁷ it is possible to develop a central repository for
26 information relating to suicide in whole population that will be of relevance internationally.
27 This paper describes the design and development of a Suicide Information Database in its
28 infancy.

29 30 31 **Authors' contributions**

32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

1
2
3
4
5
6
7 operationalisation. AJ and LK drafted the manuscript. All authors read and approved the final
8 manuscript.
9

10 11 **Funding Statement**

12
13
14 This study was funded by a grant from the National Institute for Social Care and Health
15 Research, Welsh Government, grant number RFS-12-14. DG is a National Institute for
16 Health Research (England) Senior Investigator.
17
18

19 20 **Competing interests**

21
22
23 The authors declare that they have no competing interests.
24

25 26 **Ethics approval**

27
28 Ethical approval has been granted by the Health Information Research Unit Information
29 Governance Review Panel at the College of Medicine at Swansea University.
30
31

32 33 **Data sharing statement**

34
35 It will be possible to access the data after the publication of the results. Researchers interested
36 in collaborations or further information are invited to contact AJ at a.john@swansea.ac.uk.
37
38

39 40 **References**

- 41
42
43 1. World Health Organisation. 66th World Health Assembly. Draft comprehensive
44 mental health action plan 2013-2020. A66/10 Rev.1, WHO 16 May 2013.
45 http://apps.who.int/gb/e/e_wha66.html Accessed September 2014.
46
47
48 2. Linsley KR, Schapira K, Kelly TP. Open verdict v. suicide—importance to research.
49 *Br J Psychiatry* 2001, **178**:465-468.
50
51
52
53
54

- 1
2
3
4
5
6
7 3. Appleby L, Kapur N, Shaw J, *et al*. Annual Report: England, Northern Ireland,
8 Scotland and Wales. Manchester: University of Manchester 2013.
9 [http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/Ann](http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/AnnualReport2013_UK.pdf)
10 [ualReport2013_UK.pdf](http://www.bbmh.manchester.ac.uk/cmhr/centreforsuicideprevention/nci/reports/AnnualReport2013_UK.pdf) Accessed July 2014.
11
12
- 13
14 4. Andersen UA, Andersen M, Rosholm JU, *et al*. Contacts to the health care system
15 prior to suicide: a comprehensive analysis using registers for general and psychiatric
16 hospital admissions, contacts to general practitioners and practising specialists and
17 drug prescriptions. *Acta Psychiatr Scand* 2000, **102**:126-34.
18
19
- 20 5. Qin P, Agerbo E, Mortensen PB. Suicide risk in relation to socioeconomic,
21 demographic, psychiatric, and familial factors: a national register-based study of all
22 suicides in Denmark, 1981–1997. *Am J Psychiatry* 2003, **160**:765-772.
23
24
- 25 6. Gunnell D, Löfving S, Gustafsson JE, *et al*. School performance and risk of suicide in
26 early adulthood: follow-up of two national cohorts of Swedish schoolchildren.
27 *Journal of affective disorders*, 2011, **131**:104-112.
28
29
- 30 7. Qin P, Hawton K, Mortensen PB, *et al*. Combined effects of physical illness and
31 comorbid psychiatric disorder on risk of suicide in a national population study. *Br J*
32 *Psychiatry* 2014, **204**:430-435.
33
34
- 35 8. Christiansen E, Larsen KJ, Agerbo E, *et al*. Incidence and risk factors for suicide
36 attempts in a general population of young people: a Danish register-based study. *Aust*
37 *N Zealand J Psychiatry* 2012, **47**: 259-270.
38
39
- 40 9. Björkenstam C, Johansson LA, Nordström P, *et al*. Suicide or undetermined intent? A
41 register-based study of signs of misclassification. *Population health metrics*, 2014,
42 **12**:11.
43
44
45
46
47
48
49
50
51
52
53
54

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
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.

- 1
2
3
4
5
6
7 18. Ford DV, Jones KH, Verplancke JP, *et al.* The SAIL Databank: building a national
8
9 architecture for e-health research and evaluation. *BMC Health Serv Res* 2009, **9**:1.
10
- 11 19. World Health Organization: *The ICD-10 classification of mental and behavioural*
12
13 *disorders: Clinical descriptions and diagnostic guidelines*. WHO: Geneva; 1993.
14
- 15 20. Office of National Statistics. 2011 Census Results for Wales. ONS 16th July 2012.
16
17 <http://wales.gov.uk/docs/statistics/2012/120716-census-2011-first-results-wales->
18
19 [en.pdf](http://wales.gov.uk/docs/statistics/2012/120716-census-2011-first-results-wales-en.pdf) Accessed September 2014.
20
- 21 21. Office of National Statistics. Statistical bulletin: Suicide rates in the United Kingdom,
22
23 2006 to 2010. ONS 26th January 2012. [http://www.ons.gov.uk/ons/rel/subnational-](http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2010/stb-statistical-bulletin.html)
24
25 [health4/suicides-in-the-united-kingdom/2010/stb-statistical-bulletin.html](http://www.ons.gov.uk/ons/rel/subnational-health4/suicides-in-the-united-kingdom/2010/stb-statistical-bulletin.html) Accessed
26
27 August 2014.
28
- 29 22. McGregor JI, Brooks CJ, Chalasani P, *et al.* The Health Informatics Trial
30
31 Enhancement Project (HITE): Using routinely collected primary care data to identify
32
33 potential participants for a depression trial. *Trials* 2010, **11**:39.
34
35
- 36 23. Brooks CJ, Stephens JW, Price DE, *et al.* Use of a patient linked data warehouse to
37
38 facilitate diabetes trial recruitment from primary care. *Primary Care Diabetes* 2009,
39
40 **3**:245-248
41
42
- 43 24. Hill RA, Brophy S, Brunt H, *et al.* Protocol of the baseline assessment for the
44
45 Environments for Healthy Living (EHL) Wales cohort study. *BMC Public Health*
46
47 2010, **10**:150
48
49
50
51
52
53
54

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
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.

- 1
2
3
4
5
6
7 34. Grimes DA, Schultz KF. Compared to what? Finding controls in a case control study.
8 *Lancet* 2005, **365**: 1429–33.
9
- 10
11 35. Office for National Statistics. Welsh index of Multiple Deprivation 2011. Summary
12 report, Welsh Government; 2011.
13
- 14 36. Townsend P, Phillimore P, Beattie A. *Health and Deprivation: Inequality and the*
15 *North*. Routledge: London; 1988.
16
- 17 37. Office for National Statistics. Super Output Areas 2011.
18 [http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-](http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html)
19 [output-areas--soas-/index.html](http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html) Accessed August 2014.
20
21
22
- 23 38. Bruzzi P, Green SB, Byar DP, *et al*. Estimating the population attributable risk for
24 multiple risk factors using case-control data. *Am J Epidemiol* 1985, **122**:904-914.
25
26
- 27 39. Great Britain: Data Protection Act 1998. London: HMSO.
28 <http://www.legislation.gov.uk/ukpga/1998/29/contents> Accessed August 2014.
29
30
31
- 32 40. MRC: Policy and Guidance on Sharing of Research Data from Population and Patient
33 Studies. MRC 23rd November 2011.
34 <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC008302> Accessed
35 August 2014.
36
37
38
- 39 41. Wellcome Trust: Data sharing policy. Wellcome Trust August 2010.
40 [http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-](http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm)
41 [statements/WTX035043.htm](http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm) Accessed August 2014.
42
43
44
45
46
- 47 42. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic
48 patient record data in primary care. *BMJ* 2003, **326**:1070.
49
50
51
52
53
54

- 1
2
3
4
5
6
7 43. University of Bristol, Suicide Research Centre.
8 <http://www.bristol.ac.uk/psychiatry/research/suicideresearch.html> Accessed August
9 2014.
10
11
12
13 44. The University of Manchester Centre for Suicide Prevention.
14 <http://www.bbmh.manchester.ac.uk/cmhr/research/centreforsuicideprevention/>
15 Accessed August 2014.
16
17
18
19 45. University of Oxford, Centre for Suicide Research. <http://cebmh.warne.ox.ac.uk/csr/>
20 Accessed August 2014.
21
22
23
24 46. ISD Scotland. Scottish Suicide Information Database (ScotSID). ISD Scotland 2011.
25 <http://data.gov.uk/dataset/scottish-suicide-information-database> Accessed August
26 2014.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —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 <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants	209-282
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —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) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed	236-282

1
2 *Case-control study*—If applicable, explain how matching of cases
3 and controls was addressed

4 *Cross-sectional study*—If applicable, describe analytical methods
5 taking account of sampling strategy
6

7 (e) Describe any sensitivity analyses

8 Continued on next page
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

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) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	Will
		<i>Cross-sectional study</i> —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.