Myocardial infarction and stroke subsequent to urinary tract infection (MISSOURI): protocol for a self-controlled case series using linked electronic health records

Nicola F Reeve 1, Victoria Best 2, David Gillespie 3, Kathryn Hughes 4, Fiona V Lugg-Widger 5, Rebecca Cannings-John 6, Fatemeh Torabi 2, Mandy Wootton 5, Ashley Akbari 7, Haroon Ahmed 1

ABSTRACT

Introduction There is increasing interest in the relationship between acute infections and acute cardiovascular events. Most previous research has focused on understanding whether the risk of acute cardiovascular events increases following a respiratory tract infection. The relationship between urinary tract infections (UTIs) and acute cardiovascular events is less well studied. Therefore, the aim of this study is to determine whether there is a causal relationship between UTI and acute myocardial infarction (MI) or stroke.

Methods and analysis We will undertake a self-controlled case series study using linked anonymised general practice, hospital admission and microbiology data held within the Secure Anonymised Information Linkage (SAIL) Databank. Self-controlled case series is a relatively novel study design where individuals act as their own controls, thereby inherently controlling for time-invariant confounders. Only individuals who experience an exposure and outcome of interest are included.

We will identify individuals in the SAIL Databank who have a hospital admission record for acute MI or stroke during the study period of 2010–2020. Individuals will need to be aged 30–100 during the study period and be Welsh residents for inclusion. UTI will be identified using general practice, microbiology and hospital admissions data. We will calculate the incidence of MI and stroke in predefined risk periods following an UTI and in ‘baseline’ periods (without UTI exposure) and use conditional Poisson regression models to derive incidence rate ratios.

Ethics and dissemination Data access, research permissions and approvals have been obtained from the SAIL independent Information Governance Review Panel, project number 0972. Findings will be disseminated through conferences, blogs, social media threads and peer-reviewed journals. Results will be of interest internationally to primary and secondary care clinicians who manage UTIs and may inform future clinical trials of preventative therapy.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The self-controlled case series method controls for time-invariant confounding, enabling more reliable causal estimates of the association between urinary tract infection (UTI) and acute myocardial infarction (MI) or stroke, compared with between-individual study designs.

⇒ A causal relationship between UTI and acute MI or stroke has implications for our understanding of cardiovascular disease mechanisms and may inform new methods of disease prevention.

⇒ Using individual-level population-scale anonymised, routinely collected electronic health record (EHR) data provides adequate power to study subgroups and maximises representativeness and generalisability.

⇒ EHR data are collected and recorded for clinical purposes, and therefore, the reliability of research findings is dependent on the quality and completeness of these data.

⇒ Clinical and microbiological diagnoses of UTI are subject to caveats, and therefore, we will use several definitions of UTI that use the different data sources in the Secure Anonymised Information Linkage Databank.

INTRODUCTION

Since the late 1990s, an increasing number of observational studies have found an association between acute infections and myocardial infarction (MI).1–16 Most studies focused on respiratory tract infections (RTIs), and found an increased risk of acute MI in the 1–3 days following an RTI, with the effect size varying according to the infecting organism.2–8 For example, Kwong et al found a sixfold increase in the risk of MI in the week after influenza infection, a fourfold increase after respiratory syncytial virus and a threefold increase after...
Urinary tract infections (UTIs) can affect any part of the urinary system, including the kidneys, ureters, bladder and urethra. Most infections involve the lower urinary tract: the bladder and the urethra. UTIs are common infections, with 37% of women reporting experiencing at least one in their lifetime, and 29% experiencing more than one. UTIs are associated with considerable morbidity. The global burden of disease study 2010 estimates the disability-adjusted life-years attributable to tubulointerstitial nephritis, pyelonephritis and UTI to be 45 (95% uncertainty interval 32–55) per 100,000 population.

METHODS AND ANALYSIS
Aims and objectives
The specific objectives of this research are to:
1. Estimate incidence rate ratios (IRRs) for acute MI and stroke in the 90 days following a clinically suspected UTI, with the risk being highest in the first 3 days. However, the data analysed are almost 20 years old, and there have been no attempts to replicate the findings. Furthermore, the study defined UTI using clinical codes only, so it is unclear if the reported associations related to individuals with clinical symptoms alone, or symptoms and bacteriuria, making it difficult to interpret whether individuals had true UTIs, or whether non-specific symptoms were misdiagnosed as UTI but represented early signs of a cardiovascular event. In addition, other studies have found that roughly two-thirds of women suspected to have UTI on presentation to primary care have no evidence of UTI on microbiological culture. Furthermore, the use of clinical codes alone to define UTI may lead to bias from misclassification of the exposure.

There is, therefore, the aim of the study is to determine whether there is a causal relationship between UTI and acute MI or stroke by analysing linked general practice, hospital admission and microbiology data, from a representative sample of the Welsh population. We will use the SCCS method, which controls for time-invariant confounding, enabling us to more reliably draw causal inferences between UTI and acute MI or stroke, compared with between-individual study designs.
2. Assess the effect of different methods of UTI ascertainment on estimated rate ratios (ie, clinically suspected and microbiologically confirmed; clinically diagnosed only; clinically suspected but not supported by microbiology and UTI diagnosed and treated in hospital).

3. Investigate whether the effect of a clinically suspected and microbiologically confirmed UTI on acute MI and stroke differs according to the infecting organism.

Our primary hypothesis is that a clinically suspected and microbiologically confirmed UTI will increase the risk of acute MI or stroke in the 0–90 days posts-infection period.

Data

We will use the Secure Anonymised Information Linkage (SAIL) Databank. This is an internationally recognised trusted research environment (TRE), with robust secure storage, enabling access to anonymised, linkable, individual-level Welsh population-scale data for research, with a focus on improving population health and health services. Data within SAIL are pseudonymised and made available to approved projects and users following an application to, and approval from the independent information governance review panel (IGRP). SAIL’s storage and linkage processes ensure anonymity: first, data sources being provided to SAIL are split as per the standard split file process, with the source organisation splitting the source data into demographic data and clinical data, with a system linkage field to allow data to be rejoined later. This addresses confidentiality and disclosure issues that arise when working with health data by separating easily recognised person-based variables such as name and date of birth from clinical data, including information on diagnoses, tests and prescriptions. The demographic data are anonymised and assigned an anonymised linkage field. These split files are then joined together using the system linkage field by the SAIL team and made available to researchers following encryption.\(^\text{21-23}\)

We will use the SAIL Databank to access the following linked data: Welsh Longitudinal General Practice data (WLGP), Patient Episode Database for Wales (PEDW), Welsh Results Reporting Service (WRRS). The WLGP contains data from 84\% of general practices in Wales, consisting of longitudinal data for 2.6 million Welsh residents, representing 84\% of the population.\(^\text{24}\) Demographic data, clinical diagnoses and prescription data are included. The PEDW contains International Classification of Disease version 10 (ICD-10) coded diagnoses for admissions to any Welsh hospital.\(^\text{25}\) The WRRS contains data on all tests requested from primary and secondary care National Health Service (NHS) Wales organisations processed and analysed in NHS Wales laboratories, including requests for urine microscopy, culture and antibiotic susceptibilities.\(^\text{26}\) Data are available from the data sources at varying times based on when clinical information systems began, with data quality improving over time. As such, for our study, based on our approvals and where the data sources have consistent coverage and quality, we will be using them between 1 January 2010 and 31 December 2020.

Study design

Individuals who experience UTIs and individuals who do not can differ in unmeasured ways and hence could be sources of residual confounding. We will use the SCCS design method to deal with this issue. The SCCS method is an epidemiological study design for which individuals act as their own control so that both measured and unmeasured characteristics that vary between individuals are completely controlled.\(^\text{27}\) Only individuals who have experienced an outcome and exposure of interest are included. The SCCS method compares the incidence of an outcome during predefined risk periods with incidence during baseline periods (all times outside of risk and prerisk periods) and estimates the temporal association between a transient exposure and outcome. Time-invariant covariates (eg, sex) are inherently controlled for, and time varying covariates (eg, age) are adjusted for within the models. The method was originally developed to investigate associations between vaccinations and acute adverse events, such as aseptic meningitis,\(^\text{28-29}\) but has since been applied in a range of epidemiological
**Table 2** Definitions of UTI for primary and secondary analyses. As the UTI definitions are combinations of two or more components, the start of the risk period is defined as the date of the earliest component.

<table>
<thead>
<tr>
<th>UTI-related read code in GP data (online supplemental material B)</th>
<th>Antibiotic prescription in GP data (online supplemental material C)</th>
<th>UTI-related ICD-10 code in PEDW (online supplemental material D)</th>
<th>Urine culture results in WRRS</th>
<th>Time frame</th>
<th>Clinical scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary analysis</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes, showing bacterial growth of $\geq 1 \times 10^8$ CFU/L and WCC $\geq 1 \times 10^8$ /L</td>
<td>Three codes occur within a 7-day window</td>
</tr>
<tr>
<td>Secondary analysis 1</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes, showing mixed bacterial growth (any descriptor for ‘mixed growth’ or $&gt;3$ organisms).</td>
<td>Three codes occur within a 7-day window</td>
</tr>
<tr>
<td>Secondary analysis 2</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Same day</td>
</tr>
<tr>
<td>Secondary analysis 3</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes, showing bacterial growth of $&lt;1 \times 10^7$ CFU/L.</td>
<td>Three codes occur within a 7-day window</td>
</tr>
<tr>
<td>Secondary analysis 4</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes, showing bacterial growth of $\geq 1 \times 10^8$ CFU/L and WCC $\geq 1 \times 10^8$ /L</td>
<td>Two codes occur within a 7-day window</td>
</tr>
<tr>
<td>Secondary analysis 5</td>
<td>Yes, OR ICD-10 code</td>
<td>Yes, OR ICD-10 code</td>
<td>Yes, OR: UTI Read code AND antibiotic Read code</td>
<td>Yes, showing bacterial growth of $\geq 1 \times 10^8$ CFU/L and WCC $\geq 1 \times 10^8$ /L</td>
<td>Two/three codes occur within a 7-day window</td>
</tr>
</tbody>
</table>

GP, General Practice; ICD-10, International Classification of Disease version 10; PEDW, Patient Episode Database for Wales; UTI, urinary tract infection; WCC, white cell count; WRRS, Welsh Results Reporting Service.
The SCCS method starts with identifying individuals who have had the outcome of interest. Therefore, the source population are individuals within the SAIL Databank who have a hospital admission record for acute MI or stroke during the study period. For inclusion, individuals will need to be aged 30–100 between 1 January 2010 and 31 December 2020 and be Welsh residents. We chose a lower age bound of 30 years to reduce the chance of including MI and strokes due to congenital or other non-artherosclerotic causes. A lower age bound of 40 might reduce this chance further but would increase the chance of missing relevant events, especially given the greater burden of cardiovascular disease in Wales compared with the UK as a whole.35

Outcomes
Outcomes of interest are acute MI or stroke, as identified by ICD-10 codes from inpatient diagnoses recorded in the PEDW. A list of ICD-10 codes to be used are given in online supplemental material A.

Exposure
The exposure of interest is an UTI. The risk period is 0–90 days following an UTI. This period was chosen as previous research has shown an increase in acute MI and stroke risk in the 1–90 days following an UTI.6 There will be a prerisk period of 7 days before the risk period, to allow for the situation where an individual has an UTI for several days prior to consultation, so events in the prerisk period are not erroneously attributed to the baseline period.

Baseline periods are all times outside of the risk and prerisk periods. The study period is from 1 January 2010 to 31 December 2020. The observation period is different for each individual. It generally follows the study period but may start later for some individuals who were not Welsh residents at the start of the study period but became so later, or who turned 30 years of age sometime during the study period. Similarly, the observation period may end before the study period when individuals moved out of Wales or died.

Figure 2  Case selection process. *Individuals appear and are counted in both the stroke and MI datasets if they had both an MI and a stroke event within the study period. MI, myocardial infarction; PEDW, Patient Episode Database for Wales; SAIL, Secure Anonymised Information Linkage; WLGP, Welsh Longitudinal General Practice data; WDSD, Welsh Deographic Service Dataset.
exposed to an UTI if the following events occur within a 7-day window:

1. A General Practice (GP) record of an UTI diagnostic or symptom code.
2. A GP record of an antibiotic prescription.
3. A microbiology record of a urine sample with bacterial growth of a single organism of $\geq 1 \times 10^7$ colony-forming units (CFU) per litre and white cell count (WCC) $\geq 1 \times 10^8$ per litre. If there are two organisms grown, both must demonstrate growth of $\geq 1 \times 10^8$ CFU per litre. More than two organisms will be regarded as mixed growth and thus not supportive of an UTI diagnosis.

In a sensitivity analysis, we will widen the microbiological criteria and include all urine samples with bacterial growth of a single organism of $\geq 1 \times 10^8$ CFU per litre and white cell count $\geq 1 \times 10^8$ per litre and more than two organisms will be regarded as mixed growth.

### Table 3: Characteristics of cases of stroke and MI

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Stroke</th>
<th>MI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of males (mean (25th, 75th centiles))</td>
<td>74 (64, 82)</td>
<td>69 (59, 78)</td>
</tr>
<tr>
<td>Age of females (mean (25th, 75th centiles))</td>
<td>79 (69, 87)</td>
<td>77 (66, 85)</td>
</tr>
</tbody>
</table>

We will perform several sensitivity and subgroup analyses to assess the robustness of the findings of our primary analysis to different assumptions.
We will explore the impact of using a wider definition of MI and stroke, including ICD-10 codes for acute coronary syndromes and transient ischaemic attacks. This will include events that have potentially been missed by our main definition and assess how sensitive our findings are to the definition of MI and stroke.

We will explore the impact of widening the microbiological definition of UTI to bacterial growth of a single organism of $\geq 1 \times 10^7$ CFU per litre irrespective of WCC.

We will differentiate first-ever MI or stroke from recurrent events, and report risk estimates separately. This analysis will exclude individuals with a PEDW record of an event before the observation period, and include only those who have their first ever event during the observation period.

We will extend the prerisk period to 14 days. Some individuals may have had an UTI for several days prior to diagnosis, and so an acute MI or stroke during this time may relate to exposure but without a prerisk period, would count towards the baseline period.

We will repeat the analysis excluding individuals who died within 30 days of an event to examine the potential effect of an event dependent observation period.

We will restrict the definition of UTI to include only antibiotic prescriptions for nitrofurantoin (currently recommended first-line therapy) to explore whether the choice of antibiotic impacts the findings.

We will use interaction terms to assess the impact of specific bacterial organisms on the relationship between UTI and MI or stroke.

We will examine whether the COVID-19 pandemic may affect our findings by (1) excluding individuals whose MI or stroke occurred in 2020, and (2) including an interaction term to explore whether the association between exposure and outcome differs in 2020 vs pre-2020.

We will include an interaction term to explore whether the association between exposure and outcome differs in those with and without a history of diabetes, given its potential role as a risk factor of both UTI and MI/stroke.

Sample size and power

Our initial work has identified 51,656 individuals with acute MI and 58,146 with stroke in the SAIL Databank who meet all inclusion criteria. In the previous study by Smeeth et al., the sample size was 53,709 for acute MI, and 55,157 for stroke, where 16% of acute MI cases, and 21% of stroke cases were exposed to UTI. Based on a conservative exposure rate of 10%, and an at-risk window of 90 days, we estimated the effect size that could be reliably detected with our potential sample size, using the sample size function in the SCCS package in R. The available sample provides 90% power to detect an IRR of 1.3 at the alpha=0.05 level, which is smaller than the IRRs detected in Smeeth et al.

Patient and public involvement

We developed this research proposal in collaboration with members of the Wales Centre for Primary and Emergency Care Research Service Users group (SUPER). We have a patient and public involvement (PPI) plan and are consulting the SAIL consumer panel and SUPER regarding all stages of this research, including ongoing discussion of analysis plans, review of findings, and plans for dissemination (eg, public facing outputs). We have extended our sub-group analysis to include individuals with diabetes in response to discussions with the SAIL consumer panel members. Identifying PPI for future stages of this research is a secondary objective of our PPI plan.

Summary of cases

The individuals with an ICD-10 code for either MI or stroke were selected according to the eligibility criteria, as shown in figure 2. There are 58,146 individuals with an ICD-10 code for stroke, and 51,656 individuals with an ICD-10 code for MI, a total of 105,930 unique individuals (it is possible for an individual to be in both the stroke and the MI group). Stroke cases were 49% male, MI cases 63% male. Female stroke cases were older than males, with a median age of 79 years (25–75th centiles 69–87) compared with a median of 74 years (25–75th centiles 64–82) for males. Female MI cases were also older, with a median age of 77 (25–75th centiles 66–85) compared with 69 (25–75th centiles 59–78) for males.

The mean length of observation (study entry to study exit) of stroke cases was 2796 days for females and 2981 days for males. For MI cases, 2981 for females and 3251 for males. History of diagnoses and prescription drugs is taken from all health records available for each individual. Health data is available for a median of 13 years (25–75th centiles 8–16) prior to the study start.

Characteristics of cases, including a history of diagnoses and prescription drugs prior to an event, smoking status, Welsh Index of Multiple Deprivation version 2019 and electronic Frailty Index are given in table 3.

Ethics and dissemination

All study data will be held within the SAIL Databank, an ISO27001 certified TRE for anonymised individual-level data. Data access, research permissions and approvals have been obtained from the SAIL independent IGRP, project number 0972. Analyses will be conducted within the SAIL TRE. There are strict disclosure control processes in place. Only aggregated outputs will be approved for release to ensure individuals are not identified. Findings will be disseminated through peer-review journals and conferences. Results will be of interest internationally to primary and secondary care clinicians who manage UTIs. An association between UTI and either MI or stroke will support a future funding application for a randomised trial of preventative treatments, such as anti-platelet drugs.
26. SAIL. Welsh results reports service (WRRS), 2021. Available: https://web.wwhealthdatagateway.org/dataset/71d5761-0c55-432d-82a3-bb04407-1c8d


