BMJ Open

Long-term outcomes of urinary tract infection (UTI) in Childhood (LUCI): protocol for an electronic record-linked cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-024210
Article Type:	Protocol
Date Submitted by the Author:	15-May-2018
Complete List of Authors:	Lugg-Widger, Fiona; Cardiff University, Centre for Trials Research Angel, Lianna; Cardiff University, Centre for Trials Research Cannings John, Rebeca; Cardiff University, Centre for Trials Research Jones, Hywel; Cardiff University, National Centre for Population Health and Wellbeing Research, Division of Population Medicine, Lau, Mandy; Cardiff University, Centre for Trials Research Butler, C; University of Oxford, 3. Nuffield Department of Primary Care Health Sciences Francis, Nick A.; Cardiff University, Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales), Division of Population Medicine, School of Medicine Hay, Alastair; University of Bristol, Centre for Academic Primary Care Heginbothom, Margaret; Public Health Wales Hood, Kerenza; Cardiff University, Centre for Trials Research Paranjothy, Shantini; Cardiff University, National Centre for Population Health and Wellbeing Research, Division of Population Medicine, Vandervoort, Judith; Noah's Arc Childrens Hospital for Wales Hughes, Kathryn; Cardiff University, Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales), Division of Population Medicine, School of Medicine,
Keywords:	Children, Urinary tract infections < UROLOGY, Medical record linkage, Renal scarring, Urine sampling, PRIMARY CARE

Long-term outcomes of urinary tract infection (UTI) in Childhood (LUCI): protocol for an electronic record-linked cohort study

Corresponding author:
Fiona Lugg-Widger LuggFV@cardiff.ac.uk
Address: 4th Floor Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4YS.

Lianna Angel[1] AngelL@cardiff.ac.uk
Rebecca Cannings-John[1] CanningsRL@cardiff.ac.uk
Hywel Jones[2] JonesH75@cardiff.ac.uk
Mandy Lau[1] LauTM@cardiff.ac.uk
Christopher C Butler[3] christopher.butler@phc.ox.ac.uk
Nick Francis[4] FrancisNA@cardiff.ac.uk
Alastair D Hay[5] alastair.hay@bristol.ac.uk
Margaret Heginbothom[6] margaret.heginbothom@wales.nhs.uk
Kerenza Hood[1] HoodK1@cardiff.ac.uk
Shantini Paranjothy[2] ParanjothyS@cardiff.ac.uk
Judith Van der Voort[7] judith.vandervoort@wales.nhs.uk
Kathryn Hughes[4] HughesKA6@cardiff.ac.uk

Author Affiliations

1. Centre for Trials Research, College of Biomedical & Life Sciences, Cardiff University, Cardiff, UK.
2. National Centre for Population Health and Wellbeing Research, Division of Population Medicine, Cardiff University
3. Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Primary Care Building, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2 6GG
4. Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales), Division of Population Medicine, School of Medicine, Cardiff University
5. Centre of Academic Primary Care, NIHR School for Primary Care Research, Bristol Medical School: Population Health Sciences, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS
6. Public Health Wales, 2 Capital Quarter, Tyndall Street, Cardiff CF10 4BZ
7. Noah's Arc Childrens Hospital for Wales, Cardiff, CF14 4XW

Word Count: 6382

Keywords: Children; Urinary Tract Infection; Medical record linkage; Renal scarring; Urine sampling; Primary care.

ABSTRACT

Introduction: Current guidelines advise the prompt diagnosis and treatment of UTI in children to improve both short and longer term outcomes. However, the risk of long-term complications following childhood UTI is unclear.

UTI is relatively common but difficult to diagnose in children as symptoms are non-specific. Diagnosis requires a urine sample, but sampling is difficult and infrequent, and it is not clear if sampling should be given greater priority in primary care. The LUCI study will assess the short, medium and longer-term outcomes of childhood UTI associated with routine and systematic sampling practices.

Methods and analysis: Two datasets will be established: The first will consist of routinely collected data (Hospital, GP, Microbiology) from children born and resident in Wales, linked via the Secure Anonymised Information Linkage (SAIL) databank (an 'e-cohort'). Urine sampling in this dataset reflects normal practice 'routine sampling'. Outcomes (including renal scarring, hypertension, end-stage renal failure(ESRF), hospital admissions, GP consultations, antibiotic prescriptions) for children with at least one UTI confirmed with microbiological culture (mcUTI) or no mcUTI before the age of 5 will be compared.

The second will combine data from two prospective observational studies('DUTY' & 'EURICA') employing systematic urine sampling for children presenting to primary care with acute, undifferentiated illness, linked to routine data via SAIL (Wales) and NHS Digital (England). Outcomes (as above, plus features of mcUTI) for children with a mcUTI in this dataset, identified through systematic urine sampling, will be compared to those with a mcUTI identified through routine urine sampling (dataset 1).

Ethics and dissemination: The study protocol has been approved by NHS Wales Research Ethics Committee and the Health Research Authority's Confidentiality Advisory Group. Methods of innovative study design and findings will be disseminated through peer-review journals and conferences. Results will be of interest to clinical and policy stakeholders in the UK.

Strengths and limitations of this study:

- Chronic conditions thought to be associated with childhood UTI can take many years to develop. Historically it has been difficult to obtain long-term follow-up data on large enough numbers of children. Routine data will make long-term follow-up of childhood UTI easier.
- Using a large routine dataset (hospital, microbiology, GP) from across Wales will allow a comparison of outcomes over 5 years for children with and without microbiologically confirmed UTI (mcUTI) according to routine clinical practice; and compare outcomes in these groups with those observed in high quality research data using systematic urine sampling.
- Clarifying the association of childhood UTI with chronic conditions and assessing the impact of two different sampling strategies on mcUTI outcomes, will help to prioritise interventions to improve early diagnosis, sampling and treatment, potentially improving health outcomes and reducing NHS costs.

INTRODUCTION

Urinary tract infections (UTI) are a common cause of acute illness in children and an important contributor to hospital admissions for serious bacterial infection. [1–7] In UK primary care, UTI is the cause of approximately 6% of acute illness consultations in children less than five years old. [5,7]

Childhood UTI can lead to renal scarring. [6,8,9] Renal scarring has been linked with long-term complications including hypertension, pre-eclampsia and renal failure. [10–12] Clinical guidelines promote the prompt diagnosis and treatment of UTI as a measure to prevent renal scarring and longer-term complications. [6] It is not clear what the risk of longer-term complications are for children with UTI. A systematic review in 2010 found that the prevalence of renal scarring following first childhood UTI was 15%. [8] Most included studies

were conducted in secondary care and most required fever for inclusion in the study [8]. These are likely to represent more serious UTIs than those presenting in primary care, and this rate of renal scarring may not apply to all children presenting with UTI. The risk of long-term complications (such as hypertension and renal failure) as a result of renal scarring following childhood UTI is also not clear, with some researchers questioning this association. [6,13,14] A review, commissioned by the National Institute for Health and Clinical Excellence (NICE), concluded that 'there are no appropriate studies that accurately estimate the risks of long term complications as a result of childhood UTI'. [6]

A urine sample is necessary to confirm the presence of UTI in childhood as the presenting symptoms are non-specific and similar to those found in many common childhood illnesses. [6] Furthermore, microbiological confirmation is important as some children with UTI require invasive investigations. [6] Children who become seriously ill and who are assessed in emergency departments or admitted to hospital will usually have their urine sampled. [1] However, in primary care, where most acute illness is seen, urine is infrequently sampled. It is estimated that urine is sampled in fewer than 2% of acute illness consultations with children under five years old in the UK. [4] Studies have suggested that many cases of childhood UTI are missed in primary care. [15–17] Routine practice (urine sampling based on clinician suspicion) is likely to miss more than two thirds of UTIs in primary care. [5,7] The clinical implications of missed childhood UTIs are not known. Increasing urine sampling in primary care is likely to increase the diagnosis of childhood mcUTI, and is advocated by current guidelines, but it is not clear whether this is an appropriate strategy, whether outcomes for children would improve or to what extent it should be increased. [4–6,15,18]

Clarification is needed on two issues: First, to determine what the longer-term outcomes are following childhood UTI (including those identified in primary care), and second, to determine whether outcomes vary according to sampling practice.

In this study we will describe clinical outcomes for all children with one or more mcUTI aged less than five years old, compared to those with no mcUTI, using NHS laboratory data from

across Wales. We will examine the risk factors for being diagnosed with renal scarring following mcUTI.

We will also describe longer-term follow up of clinical outcomes (including renal scarring) for at least five years following participation in two UK prospective cohort studies of acutely ill children with systematic urine sampling in primary care, the DUTY and EURICA studies. [5,7] We will compare the outcomes of those with mcUTI identified through these studies (systematic urine sampling) with those identified through routine practice.

METHODS AND DESIGN

Research objectives

The LUCI Study will use data linkage of routinely collected datasets and data from two cohorts of participants to answer two main research questions:

Research Question 1: Through routine sampling, do children who have experienced a mcUTI aged <5 years old have worse outcomes (medium (0-5 years); longer-term (>5 years)) compared to children who have not experienced a mcUTI?

Research Question 2: Are short-term (<12 months) and medium-term (1-4 years) outcomes different for children with childhood mcUTI identified through systematic sampling compared to routine sampling (standard, clinician-led sampling)?

Study Design

This is a data linkage study comprising two overarching datasets of children. Dataset 1 will comprise routinely collected health data from children born and resident in Wales. Children in this dataset will have had urine sampled according to routine practice. Routine data will be available on all children for seven years, and longer for some (i.e. children will be followed up until the date of data abstraction). Dataset 1 will be used to answer Research Question 1.

Dataset 2 will be children who participated in the EURICA or DUTY studies. Children in this dataset had their urine systematically sampled (all children presenting with an acute illness

were asked to provide a urine sample). DUTY and EURICA children will be followed-up by linking records to routinely collected health data from England (using NHS Digital) and Wales (using SAIL). Follow-up will be for a minimum of five years, and longer where data is available. Dataset 2 will be used to describe longer term follow-up for DUTY and EURICA study children. Those with mcUTI in Dataset 2 will be compared to those with mcUTI identified in Dataset 1 to answer Research Question 2.

The study formally started in October 2016 and will report to funder in October 2018. A summary of the data sources is provided in Table 1.

159 Table 1: Sources of Data

Data Provider	Data source	Dates available from - to	Indicative / key data items	Dataset	
				Dataset 1 (routine sampling)	Dataset 2 (systematic sampling)
SAIL (Wales)	GP	Jan 1994 – Oct 2016	Secondary outcomes including antibiotic prescriptions, GP consultations, chronic kidney disease, hypertension	✓	✓
	Patient Episode Database for Wales (PEDW)	Jan 1994 – April 2017	Primary & secondary outcomes & covariates including renal scarring, hospital admission, end-stage renal failure, VUR, renal/bladder surgery	✓	✓
	Welsh Demographic Service (WDS)	Jan 1994 - April 2017	Demographics	✓	✓
	Welsh Electronic Cohort of Children (WECC)	Jan 1994 - Sept 2011	Defines a child as “born in Wales”	✓	✓
	Datastore (data repository storing microbiology culture data from the Laboratory Information Management Systems in Wales)	2005 - 2014	Defines a microbiologically confirmed UTI	✓	✓
	Outpatient data	Jan 1994 - April 2017	Primary outcome – renal scarring	✓	✓
Individual Health Boards	Radiology Data	Jan 1994 – Jan 2017	Validation of the primary outcome (renal scarring)	✓	
NHS Digital (England)	Admitted Patient Care	April 2008 - Mar 2017	Primary & secondary outcomes & covariates including renal scarring, hospital admission, end-stage renal failure, VUR renal/bladder surgery		✓
	Outpatient	April 2008 - Mar 2017	Primary outcome – renal scarring		✓

Bristol University (England and Wales)	DUTY	April 2010 – April 2012	Baseline characteristics, UTI status, presenting symptoms & signs, initial clinical management		✓
Cardiff University (Wales)	EURICA	July 2008 – July 2010	Baseline characteristics, UTI status, presenting symptoms & signs, initial clinical management		✓

160

Data providers and datasets

The EURICA and DUTY Studies

This work builds on two large cohort studies of acutely ill children, aged less than five years old, presenting in primary care, in which mcUTI status was determined using systematic urine sampling [5,7]. In both studies, clinical and demographic data were collected and urine samples requested from all children included in the study and analysed in NHS microbiology laboratories. In the EURICA study, 1003 children were recruited from GP practices in Wales between March 2008 and July 2010. 5.9% children had mcUTI. Participants had a GP notes review at 6 months. In the DUTY study, 7163 children were recruited from GP practices in England and Wales between April 2010 and April 2012 and 5.6% children had mcUTI confirmed in NHS labs. An algorithm to predict UTI based on presenting symptoms and signs was developed. Neither study had sufficient follow-up to determine whether renal investigations to look for renal scarring had been undertaken or found. EURICA was funded by NISCHR (now Health and Care Research Wales) and sponsored by Cardiff University and DUTY was NIHR HTA funded and sponsored by Bristol University.

SAIL Databank

The study will use the Secure Anonymised Information Linkage (SAIL) Databank to access routinely collected data outlined in Table 1. SAIL is a repository for a broad range of routinely collected health and population data in Wales. SAIL will also act as a data safe haven for the clinical DUTY and EURICA datasets and data made available from NHS Digital and Individual Health Boards. All data will be accessed via the SAIL Gateway following Information Governance Review Panel (IGRP) approval. SAIL Databank does not handle any identifiable data therefore all data will be anonymised including data transferred from other information centres [19–21].

NHS Digital

NHS Digital is the trading name of the Health and Social Care Information Centre (HSCIC). This study will access Hospital Episode Statistics (HES) data for participants of the DUTY and EURICA study. All available Inpatient and Outpatient records belonging to each study participant will be requested and approved by the Independent Group Advising on the Release of Data (IGARD) Panel. The data requested includes diagnosis, procedures and length of episode according to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems [ICD-10] codes.

Public Health Wales

Public Health Wales will provide a data extract of urine microbiology culture results from all microbiology laboratories in Wales (Datastore) for use with this project. This will be transferred to SAIL.

Individual Health Boards

Health boards in Wales will be approached to access anonymised radiology data for patients in dataset 1. A one off data extract of patient-level attendance data for patients born between 01/01/1994 and 31/12/2012 that attended radiology between 1994 and 2016 will be transferred to SAIL. Data extracted includes examination performed, attendance data and the radiology report.

Opportunity to opt-out (dataset two)

Dataset one uses routinely collected data that is fully anonymised so we do not require individual consent in order to access these data. Dataset two involves participants from the DUTY and EURICA study and therefore requires section 251 (s251) support of the 2006

NHS Act approval from the Health Research Authority's Confidentiality Advisory Group (HRA CAG) to pass identifiable participant data legally held by Cardiff and Bristol University (as sponsors of the studies) to the data providers. This is using an opt-out/dissent model instead of obtaining further consent. In order to provide the opportunity for participants to dissent a letter has been developed to be sent to all parents of participants. Participants have the opportunity to contact the study team through an online web-form, by email, text or telephone (details provided on website) and register their dissent. These participants who register their dissent will not be included in any of the datasets sent to the information centres and therefore will not appear in the datasets for analysis. A participant representative was consulted on the layout, wording and level of information contained in the participant opt-out letter and on the study website. A key consideration was to communicate the data transfer process. The final letter was approved by both an NHS Research Ethics Committee and CAG committee as part of overall governance approval for the study.

Data Matching

For included data sets, data will be sent to NHS Digital and SAIL (via NWIS their trusted third party (TTP)) for matching using a combination of NHS number, name, address and date of birth. Matching with NHS Digital data will be by exact matching on NHS Number, Date of Birth and Postcode. This has been conducted in other studies and achieved a high match rate (99.6%) [22]. NWIS match using NHS Number, Date of Birth, Sex, Postcode. Data matching by NHS Digital will be separate to matching conducted by NWIS on behalf of SAIL. We would expect only a small number of participants matched to both English and Welsh NHS records however there is the possibility of this for those using services across the border to their current address.

The anonymised dataset

For DUTY and EURICA participants NHS Digital will retain the unique study ID assigned by the DUTY/EURICA study, with the HES records after the matching process before it is sent to SAIL Databank. This ID will be retained for data transfer to ensure the data can be linked. The same applies to data sent to NWIS. Requested data will then be linked for each participant via SAIL using the unique study ID. The DUTY and EURICA clinical datasets will be transferred to SAIL following a process of de-identification. The data flow is shown in Figure 1. The study ID will be replaced by an anonymised linking field (ALF) to enable all datasets to be linked at an individual level. The resulting dataset will contain clinical variables (from DUTY, EURICA, SAIL and NHS Digital), de-identified demographic variables (e.g. week of birth and lower super output area - LSOA) and the ALF. The key between study ID and ALF will be retained and encrypted as a further safeguard. The ALF is further encrypted (ALF-E) so that datasets and records cannot be linked across multiple projects a researcher has access to.

Study Participants

A flow chart of participants in the study is shown in Figure 2. Dataset 1 (routine sampling) will be children born and resident in Wales for the first 5 years of life; who were less than 5 years old between 1999 and 2012. The main analysis will be on children born between 01/01/2005 and 31/12/2009 to ensure that all of the first five years of life are covered by the dates when Datastore is available.

Dataset 2 (systematic sampling from DUTY and EURICA) will be children who participated in the EURICA or DUTY studies, who were not withdrawn from the study and when provided with the opportunity to opt-out, remained in the study. For research question 2 when comparing children with a mcUTI from Dataset 1 and 2, children in Group 1 will be selected to match the DUTY and EURICA study eligibility criteria as closely as possible within the constraints of using routine data. Datasets 1 and 2 will be limited to children with a mcUTI associated with a GP consultation between March 2008 and April 2012. For both datasets,

the child's mcUTI, associated with a GP consultation (defined as within 14 days prior to the sample) will be identified and defined as the index consultation. To limit the potential transfer of GP systematic sampling behaviour, children from Group 1 with index consultations between 2008 and 2012 at practices which participated in the EURICA study will be excluded as will children with index consultations between 2010 and 2012 at practices which participated in the DUTY study. Children will only be included once in each study period (i.e. a child with a sample sent within the EURICA study period could also appear in the DUTY study period). In addition, we will apply the DUTY study exclusion criteria (where possible) to Dataset 1 to make them directly comparable with Dataset 2; therefore children will be excluded from the routine sampling cohort if any criterion met:

- Known neurogenic bladder (e.g. spina bifida) or previous bladder surgery
- Prescribed antibiotics in the 7 days prior to presentation
- Taking immunosuppressant medication
- Using urinary catheters

Children with a prior history of UTI or vesicoureteric reflux (VUR) will not be excluded (and were not excluded from the DUTY or EURICA studies) and we will explore the impact of these risk factors on outcomes.

Exposure

UTI cases will be based on NHS laboratory results. For Dataset 1, this will be through microbiological culture data downloaded from Datastore. These data represent samples which have been classified as positive or negative by NHS laboratories according to their standard operating procedures. For Dataset 2, we will use the results of microbiological culture from NHS laboratories collected during the DUTY and EURICA studies as some

participants were from England (Datastore is Wales only). For Dataset 2, the presence of significant bacteriuria (pure or predominant growth of 100,000 cfu/ml of urine) will be used to define UTI.

Dataset 1 will be divided into three groups based on the first five years of life (Figure 1).

Group 1: children with at least one mcUTI

Group 2: children with at least one urine sample but no mcUTI

Group 3: children with no urine samples

Exposure is a discrete time-varying covariate and will be taken at the point of outcome; otherwise the exposure status of the child at their 5th birthday will be taken. For the main analyses, Groups 2 and 3 will be considered together as having no microbiologically confirmed UTI. Dataset 2 will similarly be divided into three groups based on their index consultation when recruited into the DUTY and EURICA studies:

Group 4: children with a mcUTI

Group 5: children who had a urine sample but no mcUTI

Group 6: children who had no urine sample

Study variables

Table 2 shows a breakdown of the baseline data and possible covariates available for children and maternal characteristics from the data collection forms for EURICA and DUTY and WDS, WECC, and for a subset with GP records. The study outcomes are summarised in Table 3.

Table 2. Child and maternal characteristics

Risk factor	**Dataset 1: Routine sampling: Source**	**Dataset 2: Systematic sampling: Source**
Age of child at first urine sample (group 1/2 only) or index consultation (Dataset 2)	Datastore	EURICA & DUTY study data
Gender	WECC	EURICA & DUTY study data
Ethnicity	WECC	WECC & DUTY study data
Deprivation quintile at birth (taken from postcode at birth)	WDS: WIMD & Townsend score	Townsend score; EURICA & DUTY study data
Maternal age at birth (years) (category)	WECC	EURICA & DUTY Welsh participants only: WECC
Birth weight (grams) (category)	WECC	EURICA & DUTY Welsh participants only: WECC
Gestational age at birth (weeks) (category)	WECC	EURICA & DUTY Welsh participants only: WECC
Ever breastfed	WECC	EURICA & DUTY study data
Uropathogen (Enterobacteriaceae or other) (mcUTI only)	Datastore	EURICA & DUTY study data
Antimicrobial resistance (mcUTI only)	Datastore	EURICA & DUTY study data
Congenital malformations (con mals)		
known to be associated with UTI/renal scarring:	PEDW	
i. Spina bifida/neuro bladder		N/A: DUTY exclusion
ii. Renal/urinary system con mals including vesico-uretero-renal reflux (VUR)		EURICA & DUTY study data
possibly associated with UTI/renal scarring:		
i. Downs Syndrome	PEDW; WECC	EURICA & DUTY study data
ii. Cerebral palsy/other paralytic syndromes	WECC	EURICA & DUTY study data
Congenital malformations **-** Major/Minor	WECC	EURICA & DUTY Welsh participants only: WECC
Comorbidities		
i. Diabetes diagnosed under the age of 5 years	PEDW	EURICA & DUTY study data*
ii. Renal or urogenital surgery	PEDW	EURICA study data* DUTY exclusion
iii. Cancer	PEDW	EURICA & DUTY study data*
iv. Immunosuppressive disease	PEDW	EURICA & DUTY study data*
v. Circumcision (aged <5 years)	PEDW; GP	EURICA & DUTY study data*
Factors for follow-up of study participants & Research Question 2		
Symptoms & signs at index consultation	-	EURICA & DUTY study data
Management at index consultation	GP	EURICA & DUTY study data
Antenatal ultrasound urinary system abnormalities	-	EURICA & DUTY study data
Family history of UTI/urinary system problems	-	EURICA & DUTY study data
Recent antibiotics (7 days prior to index consultation)	GP	EURICA & DUTY* study data

* at time of index consultation

Table 3. Study Outcomes

	Data source		
	PEDW (All Wales)	**GP**	**Datastore (All Wales)**
Primary outcome			
Renal scarring	✓		
Sensitivity analyses			
Any renal pathology codes	✓		
GP renal scarring codes		✓	
Secondary outcomes			
Hospital admissions	✓		
Day cases	✓		
Renal/urological surgery	✓		
Hypertension	✓	✓	
Chronic kidney disease	✓	✓	
Renal failure	✓		
UTIs	✓	✓	
Renal imaging		✓	
GP consultations		✓	
Antibiotics		✓	
Dysfunctional voiding		✓	
Microbiologically confirmed UTI (5-7yrs follow up)			✓

Follow-up

Follow-up of all children in Datasets 1 and 2 will continue until the first occurrence of any of: outcome, migration, or death; and for the sub-analysis of GP data, if the patient leaves the GP practice linked to SAIL or the last data collection from the general practice. For the analysis of Research Question 1, using only children whose whole first five years of life were covered by Datastore and limiting DOBs to 1/1/2005 to 31/12/2009, will mean that shortest follow-up period will be 7 years. Follow-up will be longer where data is available. For Research Question 2, short-term (<1 year), medium-term (1-5 years) and long-term (>5 years) outcomes will be examined.

Analysis

Sample size

Comparison of renal scarring in children routinely sampled for UTI with and without mcUTI: The sample size is based on the outcome of renal scarring of children with and without

mcUTI and taken as 15% [8]. Using an odds ratio of 1.5 (a conservative effect size of 5% difference between groups), 2-sided alpha 0.05, power=90%, and given a ratio of 32:1 (children diagnosed with mc UTI: children with no UTI diagnosis) in the dataset, we require 15,519 children (452 with mcUTI:15,067 without mcUTI) for analysis. Initial examination of the SAIL dataset identified just under 13,000 children less than five years old with UTI between 1999 and 2012. As our sample size calculation requires 452 with UTI, we can be confident of adequate power for this study. However, the true proportion with renal scarring is likely to be less than that reported by Shaikh [8]. Assuming a proportion of 8% with renal scarring in the mcUTI group, and an odds ratio of 1.5 (2.5% difference), 2-sided alpha=0.05, power=90%, and a ratio of 32:1, 35,833 children (1,044 with UTI: 34,789 without UTI) are required for analysis, which is still achievable.

Comparison of systematically versus routinely sampled UTI: This sample size is constrained by the number of children with a systematically sampled microbiologically confirmed UTI by NHS lab (N=374). Using the lower estimate of renal scarring of 8% of routinely sampled UTIs, then a 5% difference (13% in the systematically sampled group) would give 89% power, with a 2-sided alpha=0.05 and a ratio of 32:1 (routinely sampled UTIs 11,968: systematically sampled UTIs 374).

Both analyses use multivariable regression. Using Green's [23] formulae, assuming medium effect sizes (odds ratio=1.5), a 2-sided alpha=0.05, and power=80%, to test 20 predictors in the multivariable regression model we require at least 159 children in total. This suggests that we will be adequately powered for both analyses (given these assumptions) to examine predictors of short and medium-term outcomes.

Statistical analysis

Dataset 1: Routine sampling of UTI

Research question 1: Comparison of medium (up to 5 years) and longer term (≥5 years) outcomes in children with mcUTI versus those without mcUTI in routine sampling.

Baseline variables will be described using appropriate descriptive summaries (N (%), mean (SD), median (interquartile range)) to summarise the population for the main analyses by group 1, 2 and 3 and any marked imbalance between the groups will be identified. There will be no formal testing of between-group differences for any variables at baseline. The main comparative analyses will be carried out at a child level since outcomes relate to an individual's exposure to one or more UTI and will test the null hypothesis that there are no differences in outcomes in the first 5 years and between the age of 5 and 7 years. The primary analysis will consist of a comparison of rate of diagnosis of renal scarring (no renal scarring, renal scarring recorded <5 years, renal scarring recorded 5-7 years) in children with mcUTI versus children with no mcUTI, using a multinomial regression model. Results will be reported as relative risk ratios alongside 95% confidence intervals (CIs). We will adjust for direct covariates of renal scarring and explore the impact of indirect effects such as a mcUTI using a causal directed acyclic graph (DAG). Unadjusted and adjusted relative risk ratios will be estimated, together with 95% CIs. Cox regression will also be performed to model time to first renal scarring diagnosis to allow us to look for this outcome using all available follow-up (at least 7 years). We will estimate hazard ratios with 95% CIs for each exposure group.

Several sensitivity analyses are proposed: The primary outcome will be expanded to include any renal pathology codes due to uncertainty around whether the renal scarring codes are sufficiently sensitive to pick up all cases of renal scarring. In addition we will explore using Read codes from GP notes to pick up additional renal scarring diagnoses. Two effect modifiers were identified as a basis for sub-group analyses for the primary outcome: gender of child and presence of any renal/urological congenital anomalies. These pre-planned analyses will be conducted by the inclusion of appropriate interaction terms in the models.

Secondary outcomes will be analysed using multinomial and time to event models (Table 3). Poisson regression models will be used where the outcome is a count of event (e.g. hospital admissions, GP consultations, antibiotics prescribed); results will be represented as incidence rate ratios alongside 95% CIs. Additionally we will identify risk factors for renal scarring (as above, plus age at first mcUTI, bacterial type & resistance) in children presenting with a childhood mcUTI. In a sub sample of children linked to GP data, the 14 days prior to the urine sample submission date will be examined to determine whether there was an associated GP consultation. We will then estimate the likely rate of GP associated mcUTIs and the rate of mcUTIs submitted from another setting (such as out of hours, outpatients or as a result of a hospital admission). An identical analysis to the primary outcome will be taken to examine whether the risk of renal scarring differs between those with a mcUTI or not, between different settings and the interaction between the two. Where numbers allow, variation in outcome will be accounted for at the level of the general practice in a multilevel model. We will also describe the levels of urine sampling and incidence of mcUTI from General Practice in the routine data.

Dataset 2: Systematic sampling of UTI

Detailed study data for DUTY and EURICA participants is available including age, gender and deprivation, presenting features, GP diagnosis and acute management. Recruited children are already grouped into mcUTI or no mcUTI through the study microbiology data and outcomes will be compared according to these groups. Five year in-patient hospital outcomes will be available for all EURICA and DUTY children (the last participant of DUTY study was recruited in April 2012). We will be able to describe serious short-term (less than 1 year) and medium-term (1-5 years) outcomes, including hospital admission, renal imaging, renal scarring, VUR and renal failure outcomes for all children in this group using PEDW data in Wales and NHS Digital hospital data in England. We will be able to describe other outcomes, such as GP consultations, recurrent UTI, dysfunctional voiding, antibiotic prescriptions, hypertension, chronic kidney disease, using GP data on a sub-set. In Wales,

Datastore will also be used to look at the urine culture results and organism resistance profile for subsequent UTIs.

We will describe GP diagnosis from study data versus Read codes and acute management from the routine data in GP records for this cohort for later comparisons and also to explore the validity of using routinely collected data in these cases. We will also assess the validity of using Read codes to diagnose UTI against microbiological culture results and agreement will be measured using the Kappa statistic.

Research question 2: Comparison of short- and medium-term outcomes in children with mcUTI: routine versus systematic sampling.

We will compare the outcomes in children with mcUTI identified through routine versus systematic sampling. Children's characteristics, presentation factors, acute management and microbiology results will be described for the groups using appropriate summary statistics. In addition, we will describe blood pressure and creatinine levels for each group if recorded and explore whether comparisons can be made.

Previously mentioned short- and medium-term outcomes will be described by the two groups of routine vs. selective sampling. Predictors of outcome will be examined as before using a multilevel multinomial regression model (no event, event <1 year, event 1-5 years) and again where numbers allowed, variation in outcome will be accounted for at the level of the general practice. Associations between covariates previously described and outcome will firstly be examined. Unadjusted and adjusted relative risk ratios will be estimated, together with 95% CIs. We will compare blood pressure and creatinine levels (where available) across the groups; we expect this data to be limited so will be exploratory.

A detailed statistical analysis plan will be written prior to database lock. The reporting and presentation of results will be in accordance with the [24–26] statements to ensure the

comprehensive reporting of our observational non-randomized evaluation of a public health intervention. SPSS and Stata will be used for all analyses [27,28].

ETHICS AND DISSEMINATION

The governance surrounding dataset one differs from dataset two. Dataset one is an anonymised dataset made available from SAIL databank with only approval required from the IGRP whereas dataset two involves the transfer of identifiable data to data providers which requires ethical approval, section 251 (s251) support, IGRP and IGARD approval. In order to obtain an unbiased sample from the DUTY and EURICA cohort an opt-out consent model is being used, this was supported by both the ethics panel and the CAG panel as justification for this model of consent.

The LUCI study will report the risk of renal scarring for children with and without childhood mcUTI across the whole of Wales. The linking of routinely collected datasets will give us a large cohort including demographic, hospital in-patient and out-patient, GP and microbiology data, allowing us to define mcUTI cases and describe outcomes for all children from both primary and secondary care. Clarifying the link between UTI, renal scarring and long-term complications will inform the management of acutely ill children in primary care, where the need for urine sampling is unclear. Determining the clinical implications of 'missed' cases of UTI through our comparison of children with mcUTI identified through routine and systematic urine sampling will also help to determine the most appropriate urine sampling strategy. This study maximises the benefits of the previously funded DUTY and EURICA cohorts, representing over 8000 acutely ill children recruited from UK primary care. Significant resources were invested by funders, patients and staff to develop these cohorts. Routine data linkage will allow us to determine longer-term outcomes for these children and to determine risks of adverse outcomes. It will also pave the way for even longer-term follow-up of cohorts of children with UTI (diagnosed both systematically and routinely) which has been identified as a high research priority by NICE.

A lay summary of the results and links to publications will be made available on the University project website and the participant. The academic outputs for this study include (i) this protocol paper, (ii) main results from research question one and (ii) main results from research question two. The findings from this study will be of interest to clinicians and policy makers and may influence the management of acutely ill children and childhood UTI.

DECLARATIONS

List of abbreviations

ALF: Anonymised linking field
ALF-E: Anonymised linking field encryption
CIs: Confidence intervals
DAG: Directed acyclic graph
DOB: Date of Birth
ESRF: End-stage renal failure
HES: Hospital Episode Statistics
HRA CAG: Health Research Authority's Confidentiality Advisory Group
HSCIC: Health and Social Care Information Centre
IGARD: Independent Group Advising on the Release of Data
IGRP: Information Governance Review Panel
LSOA: Lower super output area
mcUTI: Microbiological culture urinary tract infection
NICE: National Institute for Health and Clinical Excellence
NIHR HTA: National Institute of Health Research Health Technology Assessment
NISCHR: National Institute for Social Care and Health Research
NWIS: NHS Wales Informatics Service
PEDW: Patient Episode Database for Wales
PRIME: Primary and Emergency Care Research
SAIL: Secure Anonymised Information Linkage
TTP: Trusted third party
UTI: Urinary tract infection
VUR: Vesicoureteric reflux

WDS: Welsh Demographic Service
WECC: Welsh Electronic Cohort of Children

Ethics approval and consent to participate - Ethics approval of the study has been given by the Research Ethics Committee for Wales (16/WA/0166) and the transfer and use of identifiable data has been approved by the Health Research Authority [HRA] Confidentiality Advisory Group [CAG] (16/CAG/0114).

Consent for publication - Not Applicable

Availability of data and material - Not Applicable

Competing Interests - The authors declare that they have no competing interests

Funding - This project has been funded by the Welsh Government through Health and Care Research Wales [Project number 1068].

Authors' contributions- KHu is the chief investigator of the study. All authors have contributed to and are responsible for the final design of the study. FLW is responsible for study management. RCJ and ML are responsible for statistical planning and analysis. LA and HJ are responsible for the data management. All authors have read and approved the final manuscript.

Acknowledgements

The Centre for Trials Research receives funding from Health and Care Research Wales and Cancer Research UK. Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales) receives funding from Health and Care Research Wales. Authors are supported by The Farr Institute CIPHER, funded by Arthritis Research UK, the British Heart Foundation, Cancer Research UK, the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, the Medical Research Council, the

National Institute of Health Research, the National Institute for Social Care and Health Research (Welsh Assembly Government), the Chief Scientist Office (Scottish Government Health Directorates), and the Wellcome Trust, (MRC Grant No: MR/K006525/1) and the National Centre for Population Health & Wellbeing Research (NCPHWR).

References

1 Craig JC, Williams GJ, Jones M, *et al.* The accuracy of clinical symptoms and signs for the diagnosis of serious bacterial infection in young febrile children: prospective cohort study of 15 781 febrile illnesses. *BMJ* 2010;**340**:c1594. doi:10.1136/BMJ.C1594

2 Hoberman A, Chao HP, Keller DM, *et al.* Prevalence of urinary tract infection in febrile infants. *J Pediatr* 1993;**123**:17–23.http://www.ncbi.nlm.nih.gov/pubmed/8320616 (accessed 17 Apr 2018).

3 Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am J Med* 2002;**113 Suppl 1A**:5S–13S.http://www.ncbi.nlm.nih.gov/pubmed/12113866 (accessed 17 Apr 2018).

4 Hay AD, Birnie K, Busby J, *et al.* The Diagnosis of Urinary Tract infection in Young children (DUTY): a diagnostic prospective observational study to derive and validate a clinical algorithm for the diagnosis of urinary tract infection in children presenting to primary care with an acute illness. *Health Technol Assess (Rockv)* 2016;**20**:1–294. doi:10.3310/hta20510

5 O'Brien K, Edwards A, Hood K, *et al.* Prevalence of urinary tract infection in acutely unwell children in general practice: a prospective study with systematic urine sampling. *Br J Gen Pract* 2013;**63**:e156-64. doi:10.3399/bjgp13X663127

6 Urinary tract infection in under 16s: diagnosis and management | Guidance and guidelines | NICE. https://www.nice.org.uk/guidance/cg54 (accessed 24 Apr 2018).

7 Butler CC, O'Brien K, Pickles T, *et al.* Childhood urinary tract infection in primary care: a prospective observational study of prevalence, diagnosis, treatment, and recovery. *Br J Gen Pract* 2015;**65**:e217-23. doi:10.3399/bjgp15X684361

8 Shaikh N, Ewing AL, Bhatnagar S, *et al.* Risk of Renal Scarring in Children With a First Urinary Tract Infection: A Systematic Review. *Pediatrics* 2010;**126**:1084–91.

doi:10.1542/peds.2010-0685

9 Coulthard MG, Lambert HJ, Keir MJ. Occurrence of renal scars in children after their first referral for urinary tract infection. *BMJ* 1997;**315**:918–9.http://www.ncbi.nlm.nih.gov/pubmed/9361542 (accessed 17 Apr 2018).

10 Jacobson SH, Eklöf O, Eriksson CG, *et al.* Development of hypertension and uraemia after pyelonephritis in childhood: 27 year follow up. *BMJ* 1989;**299**:703–6.http://www.ncbi.nlm.nih.gov/pubmed/2508881 (accessed 17 Apr 2018).

11 Round J, Fitzgerald AC, Hulme C, *et al.* Urinary tract infections in children and the risk of ESRF. *Acta Paediatr* 2012;**101**:278–82. doi:10.1111/j.1651-2227.2011.02542.x

12 Smellie JM, Prescod NP, Shaw PJ, *et al.* Childhood reflux and urinary infection: a follow-up of 10-41 years in 226 adults. *Pediatr Nephrol* 1998;**12**:727–36.http://www.ncbi.nlm.nih.gov/pubmed/9874316 (accessed 17 Apr 2018).

13 Craig JC, Williams GJ. Denominators do matter: it's a myth--urinary tract infection does not cause chronic kidney disease. *Pediatrics* 2011;**128**:984–5. doi:10.1542/peds.2011-2631

14 Salo J, Ikäheimo R, Tapiainen T, *et al.* Childhood urinary tract infections as a cause of chronic kidney disease. *Pediatrics* 2011;**128**:840–7. doi:10.1542/peds.2010-3520

15 Coulthard MG, Vernon SJ, Lambert HJ, *et al.* A nurse led education and direct access service for the management of urinary tract infections in children: prospective controlled trial. *BMJ* 2003;**327**:656. doi:10.1136/bmj.327.7416.656

16 Jadresic L, Cartwright K, Cowie N, *et al.* Investigation of urinary tract infection in childhood. *BMJ* 1993;**307**:761–4. doi:10.1136/BMJ.307.6907.761

17 van der Voort J, Edwards A, Roberts R, *et al.* The struggle to diagnose UTI in children under two in primary care. *Fam Pract* 1997;**14**:44–8.http://www.ncbi.nlm.nih.gov/pubmed/9061344 (accessed 17 Apr 2018).

18 Hollingworth W, Busby J, Butler CC, *et al.* The Diagnosis of Urinary Tract Infection in Young Children (DUTY) Study Clinical Rule: Economic Evaluation. *Value Heal* 2017;**20**:556–66. doi:10.1016/j.jval.2017.01.003

19 Jones KH, Ford D V., Jones C, *et al.* A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation. *J Biomed Inform* 2014;**50**:196–204. doi:10.1016/j.jbi.2014.01.003

20 Ford D V, Jones KH, Verplancke J-P, *et al.* The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;**9**:157. doi:10.1186/1472-6963-9-157

21 Lyons RA, Jones KH, John G, *et al.* The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009;**9**:3. doi:10.1186/1472-6947-9-3

22 Robling M, Bekkers M-J, Bell K, *et al.* Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial. *Lancet* 2016;**387**:146–55. doi:10.1016/S0140-6736(15)00392-X

23 Green SB. How Many Subjects Does It Take To Do A Regression Analysis. *Multivariate Behav Res* 1991;**26**:499–510. doi:10.1207/s15327906mbr2603_7

24 Gilbert R, Lafferty R, Hagger-Johnson G, *et al.* GUILD: GUidance for Information about Linking Data sets†. *J Public Health (Bangkok)* 2018;**40**:191–8. doi:10.1093/pubmed/fdx037

25 von Elm E, Altman DG, Egger M, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;**335**:806–8. doi:10.1136/bmj.39335.541782.AD

26 Benchimol EI, Smeeth L, Guttmann A, *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Med* 2015;**12**:e1001885. doi:10.1371/journal.pmed.1001885

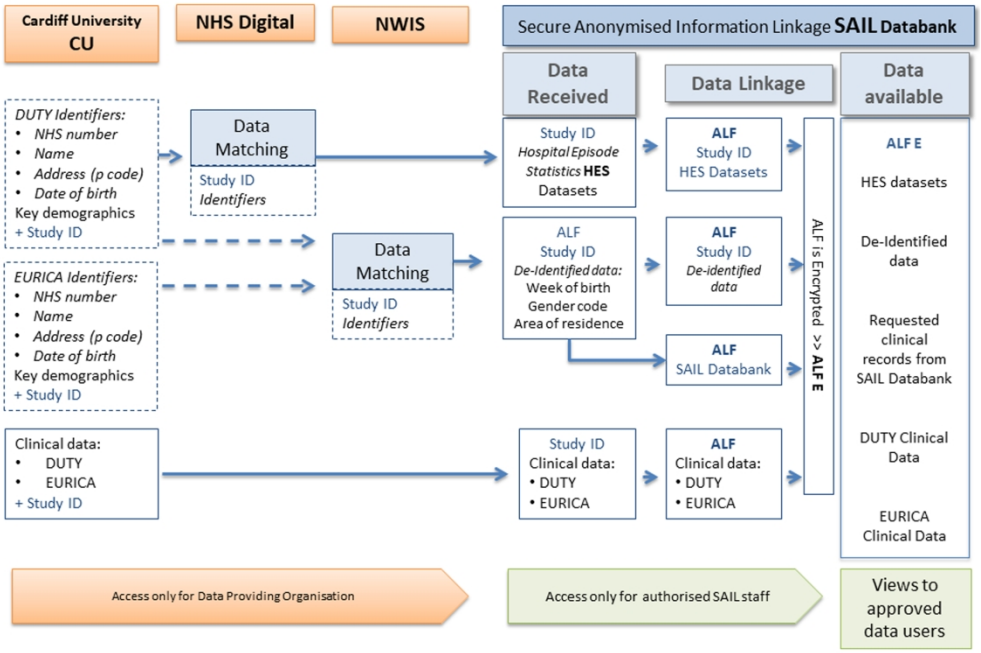
27 IBM. IBM SPSS Statistics for Windows: Version 22.0. 2013.

28 StataCorp. Stata Statistical Software: Release 15. 2017. 2017. doi:10.2307/2234838

(Title) Figure 1. The data flow for dataset 2.

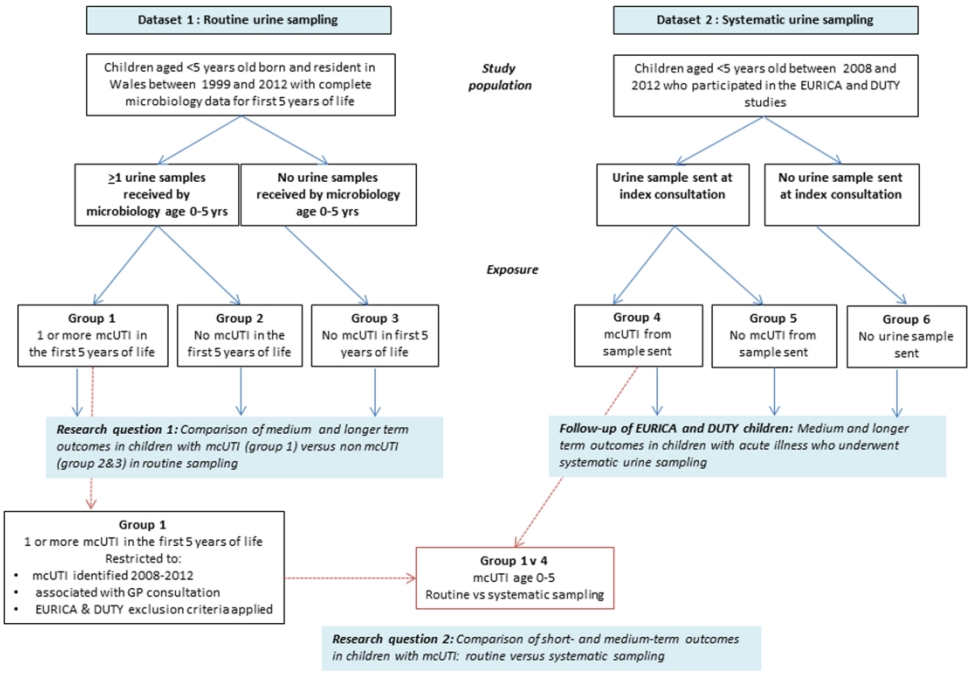
(Legend) ALF- Anonymised Linking Field; NWIS – NHS Wales Informatics Service; HES – Hospital Episode Statistics; SAIL – Secure Anonymised Information Linkage

(Title) Figure 2. Flow chart of study participants



(Title) Figure 1. The data flow for dataset 2.
(Legend) ALF- Anonymised Linking Field; NWIS – NHS Wales Informatics Service; HES – Hospital Episode Statistics; SAIL – Secure Anonymised Information Linkage

297x209mm (300 x 300 DPI)



(Title) Figure 2. Flow chart of study participants

297x209mm (300 x 300 DPI)

BMJ Open

Long-term outcomes of urinary tract infection (UTI) in Childhood (LUCI): protocol for an electronic record-linked cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-024210.R1
Article Type:	Protocol
Date Submitted by the Author:	09-Nov-2018
Complete List of Authors:	Lugg-Widger, Fiona; Cardiff University, Centre for Trials Research Angel, Lianna; Cardiff University, Centre for Trials Research Cannings-John, Rebecca; Cardiff University, Centre for Trials Research Jones, Hywel; Cardiff University, National Centre for Population Health and Wellbeing Research, Division of Population Medicine, Lau, Mandy; Cardiff University, Centre for Trials Research Butler, C; University of Oxford, 3. Nuffield Department of Primary Care Health Sciences Francis, Nick A.; Cardiff University, Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales), Division of Population Medicine, School of Medicine Hay, Alastair; University of Bristol, Centre for Academic Primary Care Heginbothom, Margaret; Public Health Wales Hood, Kerenza; Cardiff University, Centre for Trials Research Paranjothy, Shantini; Cardiff University, National Centre for Population Health and Wellbeing Research, Division of Population Medicine, Vandervoort, Judith; Noah's Arc Childrens Hospital for Wales Hughes, Kathryn; Cardiff University, Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales), Division of Population Medicine, School of Medicine,
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Urology, Paediatrics
Keywords:	Children, Urinary tract infections < UROLOGY, Medical record linkage, Renal scarring, Urine sampling, PRIMARY CARE

Long-term outcomes of urinary tract infection (UTI) in Childhood (LUCI): protocol for an electronic record-linked cohort study

Corresponding author:
Fiona Lugg-Widger LuggFV@cardiff.ac.uk
Address: 4th Floor Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4YS.

Lianna Angel[1] AngelL@cardiff.ac.uk
Rebecca Cannings-John[1] CanningsRL@cardiff.ac.uk
Hywel Jones[2] JonesH75@cardiff.ac.uk
Mandy Lau[1] LauTM@cardiff.ac.uk
Christopher C Butler[3] christopher.butler@phc.ox.ac.uk
Nick Francis[4] FrancisNA@cardiff.ac.uk
Alastair D Hay[5] alastair.hay@bristol.ac.uk
Margaret Heginbothom[6] margaret.heginbothom@wales.nhs.uk
Kerenza Hood[1] HoodK1@cardiff.ac.uk
Shantini Paranjothy[2] ParanjothyS@cardiff.ac.uk
Judith Van der Voort[7] judith.vandervoort@wales.nhs.uk
Kathryn Hughes[4] HughesKA6@cardiff.ac.uk

Author Affiliations

1. Centre for Trials Research, College of Biomedical & Life Sciences, Cardiff University, Cardiff, UK.
2. National Centre for Population Health and Wellbeing Research, Division of Population Medicine, Cardiff University
3. Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Primary Care Building, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2 6GG
4. Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales), Division of Population Medicine, School of Medicine, Cardiff University
5. Centre of Academic Primary Care, NIHR School for Primary Care Research, Bristol Medical School: Population Health Sciences, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS
6. Public Health Wales, 2 Capital Quarter, Tyndall Street, Cardiff CF10 4BZ
7. Noah's Arc Childrens Hospital for Wales, Cardiff, CF14 4XW

Word Count: 6810

Keywords: Children; Urinary Tract Infection; Medical record linkage; Renal scarring; Urine sampling; Primary care.

ABSTRACT

Introduction: Current guidelines advise the prompt diagnosis and treatment of UTI in children to improve both short and longer term outcomes. However, the risk of long-term complications following childhood UTI is unclear.

UTI is relatively common but difficult to diagnose in children as symptoms are non-specific. Diagnosis requires a urine sample, but sampling is difficult and infrequent, and it is not clear if sampling should be given greater priority in primary care. The LUCI study will assess the short, medium and longer-term outcomes of childhood UTI associated with routine and systematic sampling practices.

Methods and analysis: Two datasets will be established: The first will consist of routinely collected data (Hospital, GP, Microbiology) from children born and resident in Wales, linked via the Secure Anonymised Information Linkage (SAIL) databank (an 'e-cohort'). Urine sampling in this dataset reflects normal practice 'routine sampling'. Outcomes (including renal scarring, hypertension, end-stage renal failure(ESRF), hospital admissions, GP consultations, antibiotic prescriptions) for children with at least one UTI confirmed with microbiological culture (mcUTI) or no mcUTI before the age of 5 will be compared.

The second will combine data from two prospective observational studies('DUTY' & 'EURICA') employing systematic urine sampling for children presenting to primary care with acute, undifferentiated illness, linked to routine data via SAIL (Wales) and NHS Digital (England). Outcomes (as above, plus features of mcUTI) for children with a mcUTI in this dataset, identified through systematic urine sampling, will be compared to those with a mcUTI identified through routine urine sampling (dataset 1).

Ethics and dissemination: The study protocol has been approved by NHS Wales Research Ethics Committee and the Health Research Authority's Confidentiality Advisory Group. Methods of innovative study design and findings will be disseminated through peer-review journals and conferences. Results will be of interest to clinical and policy stakeholders in the UK.

Strengths and limitations of this study:

- Chronic conditions thought to be associated with childhood UTI can take many years to develop. Historically it has been difficult to obtain long-term follow-up data on large enough numbers of children. Routine data will make long-term follow-up of childhood UTI easier.
- Using a large routine dataset (hospital, microbiology, GP) from across Wales will allow a comparison of outcomes over 5 years for children with and without microbiologically confirmed UTI (mcUTI) according to routine clinical practice; and compare outcomes in these groups with those observed in high quality research data using systematic urine sampling.
- Clarifying the association of childhood UTI with chronic conditions and assessing the impact of two different sampling strategies on mcUTI outcomes, will help to prioritise interventions to improve early diagnosis, sampling and treatment, potentially improving health outcomes and reducing NHS costs.

INTRODUCTION

Urinary tract infections (UTI) are a common cause of acute illness in children and an important contributor to hospital admissions for serious bacterial infection. [1–7] In UK primary care, UTI is the cause of approximately 6% of acute illness consultations in children less than five years old. [5,7]

Childhood UTI can lead to renal scarring. [6,8,9] Renal scarring has been linked with long-term complications including hypertension, pre-eclampsia and renal failure. [10–12] Clinical guidelines promote the prompt diagnosis and treatment of UTI as a measure to prevent renal scarring and longer-term complications. [6] It is not clear what the risk of longer-term complications are for children with UTI. A systematic review in 2010 found that the prevalence of renal scarring following first childhood UTI was 15%. [8] Most included studies

were conducted in secondary care and most required fever for inclusion in the study [8]. These are likely to represent more serious UTIs than those presenting in primary care, and this rate of renal scarring may not apply to all children presenting with UTI. The risk of long-term complications (such as hypertension and renal failure) as a result of renal scarring following childhood UTI is also not clear, with some researchers questioning this association. [6,13,14] A review, commissioned by the National Institute for Health and Clinical Excellence (NICE), concluded that 'there are no appropriate studies that accurately estimate the risks of long term complications as a result of childhood UTI'. [6]

A urine sample is necessary to confirm the presence of UTI in childhood as the presenting symptoms are non-specific and similar to those found in many common childhood illnesses (e.g. respiratory infections, gastroenteritis, tonsillitis, ear infections) such as high temperature, poor feeding, vomiting, lethargy and irritability. [6] Furthermore, microbiological confirmation is important as some children with UTI require invasive investigations. [6] Children who become seriously ill and who are assessed in emergency departments or admitted to hospital will usually have their urine sampled. [1] However, in primary care, where most acute illness is seen, urine is infrequently sampled. It is estimated that urine is sampled in fewer than 2% of acute illness consultations with children under five years old in the UK. [4] Studies have suggested that many cases of childhood UTI are missed in primary care. [15–17] Routine practice (urine sampling based on clinician suspicion) is likely to miss more than two thirds of UTIs in primary care. [5,7] The clinical implications of missed childhood UTIs are not known. Increasing urine sampling in primary care is likely to increase the diagnosis of childhood mcUTI, and is advocated by current guidelines, but it is not clear whether this is an appropriate strategy, whether outcomes for children would improve or to what extent it should be increased. [4–6,15,18]

Clarification is needed on two issues: First, to determine what the longer-term outcomes are following childhood UTI (including those identified in primary care), and second, to determine whether outcomes vary according to sampling practice.

In this study we will describe clinical outcomes for all children with one or more mcUTI aged less than five years old, compared to those with no mcUTI, using NHS laboratory data from across Wales. We will examine the risk factors for being diagnosed with renal scarring following mcUTI.

We will also describe longer-term follow up of clinical outcomes for at least five years following participation in two UK prospective cohort studies of acutely ill children with systematic urine sampling in primary care, the DUTY and EURICA studies. [5,7] We will compare the outcomes of those with mcUTI identified through these studies (systematic urine sampling) with those identified through routine practice.

METHODS AND DESIGN

Research objectives

The LUCI Study will use data linkage of routinely collected datasets and data from two cohorts of participants to answer two main research questions:

Research Question 1: Through routine sampling, do children who have experienced a mcUTI aged <5 years old have worse outcomes (medium (0-5 years); longer-term (>5 years)) compared to children who have not experienced a mcUTI?

Research Question 2: Are short-term (<12 months) and medium-term (1-4 years) outcomes different for children with childhood mcUTI identified through systematic sampling compared to routine sampling (standard, clinician-led sampling)?

Study Design

This is a data linkage study comprising two overarching datasets of children. Dataset 1 will comprise routinely collected health data from children born and resident in Wales. Children in this dataset will have had urine sampled according to routine practice. Routine data will be available on all children for seven years, and longer for some (i.e. children will be followed up until the date of data abstraction). Dataset 1 will be used to answer Research Question 1.

Dataset 2 will be children who participated in the EURICA or DUTY studies. Children in this dataset had their urine systematically sampled (all children presenting with an acute illness were asked to provide a urine sample). DUTY and EURICA children will be followed-up by linking records to routinely collected health data from England (using NHS Digital) and Wales (using SAIL). Follow-up will be for a minimum of five years, and longer where data is available. Dataset 2 will be used to describe longer term follow-up for DUTY and EURICA study children. Those with mcUTI in Dataset 2 will be compared to those with mcUTI identified in Dataset 1 to answer Research Question 2.

The study formally started in October 2016 and will report to funder in December 2018. A summary of the data sources is provided in Table 1.

Table 1: Sources of Data

Data Provider	Data source	Dates available from - to	Indicative / key data items	Dataset	
				Dataset 1 (routine sampling)	Dataset 2 (systematic sampling)
SAIL (Wales)	GP	Jan 1994 – Oct 2016	Secondary outcomes including antibiotic prescriptions, GP consultations, chronic kidney disease, hypertension	✓	✓
	Patient Episode Database for Wales (PEDW)	Jan 1994 – April 2017	Primary & secondary outcomes & covariates including renal scarring, hospital admission, end-stage renal failure, VUR, renal/bladder surgery	✓	✓
	Welsh Demographic Service (WDS)	Jan 1994 - April 2017	Demographics	✓	✓
	Welsh Electronic Cohort of Children (WECC)	Jan 1994 - Sept 2011	Defines a child as “born in Wales”	✓	✓
	Datastore (data repository storing microbiology culture data from the Laboratory Information Management Systems in Wales)	2005 - 2014	Defines a microbiologically confirmed UTI	✓	✓
	Outpatient data	Jan 1994 - April 2017	Primary outcome – renal scarring	✓	✓
Individual Health Boards	Radiology Data	Jan 1994 – Jan 2017	Validation of the primary outcome (renal scarring)	✓	
NHS Digital (England)	Admitted Patient Care	April 2008 - Mar 2017	Primary & secondary outcomes & covariates including renal scarring, hospital admission, end-stage renal failure, VUR renal/bladder surgery		✓
	Outpatient	April 2008 - Mar 2017	Primary outcome – renal scarring		✓

Bristol University (England and Wales)	DUTY	April 2010 – April 2012	Baseline characteristics, UTI status, presenting symptoms & signs, initial clinical management		✓
Cardiff University (Wales)	EURICA	July 2008 – July 2010	Baseline characteristics, UTI status, presenting symptoms & signs, initial clinical management		✓

162

Data providers and datasets

The EURICA and DUTY Studies

This work builds on two large cohort studies of acutely ill children, aged less than five years old, presenting in primary care, in which mcUTI status was determined using systematic urine sampling [5,7]. In both studies, clinical and demographic data were collected and urine samples requested from all children included in the study and analysed in NHS microbiology laboratories. In the EURICA study, 1003 children were recruited from GP practices in Wales between March 2008 and July 2010. 5.9% children had mcUTI. Participants had a GP notes review at 6 months. In the DUTY study, 7163 children were recruited from GP practices in England and Wales between April 2010 and April 2012 and 5.6% children had mcUTI confirmed in NHS labs. An algorithm to predict UTI based on presenting symptoms and signs was developed. Neither study had sufficient follow-up to determine whether renal investigations to look for renal scarring had been undertaken or found. EURICA was funded by NISCHR (now Health and Care Research Wales) and sponsored by Cardiff University and DUTY was NIHR HTA funded and sponsored by Bristol University.

SAIL Databank

The study will use the Secure Anonymised Information Linkage (SAIL) Databank to access routinely collected data outlined in Table 1. SAIL is a repository for a broad range of routinely collected health and population data in Wales. SAIL will also act as a data safe haven for the clinical DUTY and EURICA datasets and data made available from NHS Digital and Individual Health Boards. All data will be accessed via the SAIL Gateway following Information Governance Review Panel (IGRP) approval. SAIL Databank does not handle any identifiable data therefore all data will be anonymised including data transferred from other information centres [19–21].

NHS Digital

NHS Digital is the trading name of the Health and Social Care Information Centre (HSCIC). This study will access Hospital Episode Statistics (HES) data for participants of the DUTY and EURICA study. All available Inpatient and Outpatient records belonging to each study participant will be requested and approved by the Independent Group Advising on the Release of Data (IGARD) Panel. The data requested includes diagnosis, procedures and length of episode according to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems [ICD-10] codes.

Public Health Wales

Public Health Wales will provide a data extract of urine microbiology culture results from all microbiology laboratories in Wales (Datastore) for use with this project. This will be transferred to SAIL.

Individual Health Boards

Health boards in Wales will be approached to access anonymised radiology data for patients in dataset 1. A one off data extract of patient-level attendance data for patients born between 01/01/1994 and 31/12/2012 that attended radiology between 1994 and 2016 will be transferred to SAIL. Data extracted includes examination performed, attendance data and the radiology report.

Opportunity to opt-out (dataset two)

Dataset one uses routinely collected data that is fully anonymised so we do not require individual consent in order to access these data. Dataset two involves participants from the DUTY and EURICA study and therefore requires section 251 (s251) support of the 2006

NHS Act approval from the Health Research Authority's Confidentiality Advisory Group (HRA CAG) to pass identifiable participant data legally held by Cardiff and Bristol University (as sponsors of the studies) to the data providers. This is using an opt-out/dissent model instead of obtaining further consent. In order to provide the opportunity for participants to dissent a letter has been developed to be sent to all parents of participants. Participants have the opportunity to contact the study team through an online web-form, by email, text or telephone (details provided on website) and register their dissent. These participants who register their dissent will not be included in any of the datasets sent to the information centres and therefore will not appear in the datasets for analysis. A participant representative was consulted on the layout, wording and level of information contained in the participant opt-out letter and on the study website. A key consideration was to communicate the data transfer process. The final letter was approved by both an NHS Research Ethics Committee and CAG committee as part of overall governance approval for the study.

Data Matching

For included data sets, data will be sent to NHS Digital and SAIL (via NWIS their trusted third party (TTP)) for matching using a combination of NHS number, name, address and date of birth. Matching with NHS Digital data will be by exact matching on NHS Number, Date of Birth and Postcode. This has been conducted in other studies and achieved a high match rate (99.6%) [22]. NWIS match using NHS Number, Date of Birth, Sex, Postcode. Data matching by NHS Digital will be separate to matching conducted by NWIS on behalf of SAIL. We would expect only a small number of participants matched to both English and Welsh NHS records however there is the possibility of this for those using services across the border to their current address.

The anonymised dataset

For DUTY and EURICA participants NHS Digital will retain the unique study ID assigned by the DUTY/EURICA study, with the HES records after the matching process before it is sent to SAIL Databank. This ID will be retained for data transfer to ensure the data can be linked. The same applies to data sent to NWIS. Requested data will then be linked for each participant via SAIL using the unique study ID. The DUTY and EURICA clinical datasets will be transferred to SAIL following a process of de-identification. The data flow is shown in Figure 1. The study ID will be replaced by an anonymised linking field (ALF) to enable all datasets to be linked at an individual level. The resulting dataset will contain clinical variables (from DUTY, EURICA, SAIL and NHS Digital), de-identified demographic variables (e.g. week of birth and lower super output area - LSOA) and the ALF. The key between study ID and ALF will be retained and encrypted as a further safeguard. The ALF is further encrypted (ALF-E) so that datasets and records cannot be linked across multiple projects a researcher has access to.

Study Participants

A flow chart of participants in the study is shown in Figure 2. Dataset 1 (routine sampling) will be children born and resident in Wales for the first 5 years of life; who were less than 5 years old between 1999 and 2012. The main analysis will be on children born between 01/01/2005 and 31/12/2009 to ensure that all of the first five years of life are covered by the dates when Datastore is available.

Dataset 2 (systematic sampling from DUTY and EURICA) will be children who participated in the EURICA or DUTY studies, who were not withdrawn from the study and when provided with the opportunity to opt-out, remained in the study. For research question 2 when comparing children with a mcUTI from Dataset 1 and 2, children in Group 1 will be selected to match the DUTY and EURICA study eligibility criteria as closely as possible within the constraints of using routine data. Datasets 1 and 2 will be limited to children with a mcUTI associated with a GP consultation between March 2008 and April 2012. For both datasets,

the child's mcUTI, associated with a GP consultation (defined as within 14 days prior to the sample) will be identified and defined as the index consultation. To limit the potential transfer of GP systematic sampling behaviour, children from Group 1 with index consultations between 2008 and 2012 at practices which participated in the EURICA study will be flagged as will children with index consultations between 2010 and 2012 at practices which participated in the DUTY study. Children will only be included once in each study period (i.e. a child with a sample sent within the EURICA study period could also appear in the DUTY study period). In addition, we will apply the DUTY study exclusion criteria (where possible) to Dataset 1 to make them directly comparable with Dataset 2; therefore children will be excluded from the routine sampling cohort if any criterion met:

- Known neurogenic bladder (e.g. spina bifida) or previous bladder surgery
- Prescribed antibiotics in the 7 days prior to presentation
- Taking immunosuppressant medication
- Using urinary catheters

Children with a prior history of UTI or vesicoureteric reflux (VUR) will not be excluded (and were not excluded from the DUTY or EURICA studies) and we will explore the impact of these risk factors on outcomes.

Exposure

UTI cases will be based on NHS laboratory results. For Dataset 1, this will be through microbiological culture data downloaded from Datastore. These data represent samples (from both community and hospital settings) which have been classified as positive or negative by NHS laboratories according to their standard operating procedures. We do not know how urine was sampled, and this is likely to vary between settings. In most cases,

these are likely to be clean catch samples, but may include urine collection pads or bags (particularly in community samples) as recommended by NICE; or catheter or suprapubic aspiration (SPA) in hospital samples.[6] NHS laboratories take into consideration the nature of the urine sample in their reporting.[23] For Dataset 2, we will use the results of microbiological culture from NHS laboratories collected during the DUTY and EURICA studies as some participants were from England (Datastore is Wales only). For Dataset 2, the presence of significant bacteriuria (pure or predominant growth of 100,000 cfu/ml of urine) will be used to define UTI.

For Dataset 1 we define the exposure period as <5 years and will be grouped as follows: (Figure 1).

Group 1: children with at least one mcUTI before their 5th birthday or before outcome of interest

Group 2: children with at least one urine sample but no mcUTI before their 5th birthday or before outcome of interest

Group 3: children with no urine samples before their 5th birthday or before outcome of interest

Exposure is a discrete time-varying covariate (0 (group 2 or 3 no UTI or no sample respectively) until first exposure of group 1 (mcUTI) thereafter) and each patient's exposure status will be taken at the point of each outcome; otherwise the exposure status of the child at their 5th birthday will be taken.

For the main analyses, Groups 2 and 3 will be considered together as having no microbiologically confirmed UTI. Dataset 2 will similarly be divided into three groups based on their index consultation when recruited into the DUTY and EURICA studies:

Group 4: children with a mcUTI

Group 5: children who had a urine sample but no mcUTI

Group 6: children who had no urine sample

Study variables

Table 2 shows a breakdown of the baseline data and possible covariates available for children and maternal characteristics from the data collection forms for EURICA and DUTY and WDS, WECC, and for a subset with GP records. The study outcomes are summarised in Table 3.

Table 2. Child and maternal characteristics

Risk factor	**Dataset 1: Routine sampling: Source**	**Dataset 2: Systematic sampling: Source**
Age of child at first urine sample (group 1/2 only) or index consultation (Dataset 2)	Datastore	EURICA & DUTY study data
Gender	WECC	EURICA & DUTY study data
Ethnicity	WECC	WECC & DUTY study data
Deprivation quintile at birth (taken from postcode at birth)	WDS: WIMD & Townsend score	Townsend score; EURICA & DUTY study data
Maternal age at birth (years) (category)	WECC	EURICA & DUTY Welsh participants only: WECC
Birth weight (grams) (category)	WECC	EURICA & DUTY Welsh participants only: WECC
Gestational age at birth (weeks) (category)	WECC	EURICA & DUTY Welsh participants only: WECC
Ever breastfed	WECC	EURICA & DUTY study data
Uropathogen (Enterobacteriaceae or other) (mcUTI only)	Datastore	EURICA & DUTY study data
Antimicrobial resistance (mcUTI only)	Datastore	EURICA & DUTY study data
Congenital malformations (con mals)		
known to be associated with UTI/renal scarring:	PEDW	
i. Spina bifida/neuro bladder		N/A: DUTY exclusion
ii. Renal/urinary system con mals including vesico-uretero-renal reflux (VUR)		EURICA & DUTY study data
possibly associated with UTI/renal scarring:		
i. Downs Syndrome	PEDW;	EURICA & DUTY study data

Risk factor	Dataset 1: Routine sampling: Source	Dataset 2: Systematic sampling: Source
	WECC	
ii. Cerebral palsy/other paralytic syndromes	WECC	EURICA & DUTY study data
Congenital malformations - Major/Minor	WECC	EURICA & DUTY Welsh participants only: WECC
Comorbidities		
i. Diabetes diagnosed under the age of 5 years	PEDW	EURICA & DUTY study data*
ii. Renal or urogenital surgery	PEDW	EURICA study data* DUTY exclusion
iii. Cancer	PEDW	EURICA & DUTY study data*
iv. Immunosuppressive disease	PEDW	EURICA & DUTY study data*
v. Circumcision (aged <5 years)	PEDW; GP	EURICA & DUTY study data*
Factors for follow-up of study participants & Research Question 2		
Symptoms & signs at index consultation	-	EURICA & DUTY study data
Management at index consultation	GP	EURICA & DUTY study data
Antenatal ultrasound urinary system abnormalities	-	EURICA & DUTY study data
Family history of UTI/urinary system problems	-	EURICA & DUTY study data
Recent antibiotics (7 days prior to index consultation)	GP	EURICA & DUTY* study data

* at time of index consultation

Table 3. Study Outcomes

	Data source		
	PEDW (All Wales)	**GP**	**Datastore (All Wales)**
Primary outcome			
Renal scarring	✓		
Sensitivity analyses			
Any renal pathology codes	✓		
GP renal scarring codes		✓	
Secondary outcomes			
Hospital admissions	✓		
Day cases	✓		
Renal/urological surgery	✓		
Hypertension	✓	✓	
Chronic kidney disease	✓	✓	
Renal failure	✓		
UTIs	✓	✓	
Renal imaging		✓	
GP consultations		✓	
Antibiotics		✓	
Dysfunctional voiding		✓	
Microbiologically confirmed UTI (5-7yrs follow up)			✓

Follow-up

Follow-up of all children in Datasets 1 and 2 will continue until the first occurrence of any of: outcome, migration, death or end of follow-up; and for the sub-analysis of GP data, if the patient leaves the GP practice linked to SAIL or the last data collection from the general practice. For the analysis of Research Question 1, using only children whose whole first five years of life were covered by the dates that Datastore was available (excluding children whos first five years of life fall outside of Datastore availability) and limiting DOBs to 1/1/2005 to 31/12/2009, will mean that shortest follow-up period will be 7 years. Follow-up will be longer where data is available. For Research Question 2, we will examine outcomes at 30 days and at 1 year (short term), 1-5 years (medium-term) and >5 years (long-term) after the index consultation.

Analysis

Sample size

Comparison of renal scarring in children routinely sampled for UTI with and without mcUTI: The sample size is based on the outcome of renal scarring of children with and without mcUTI and taken as 15% [8]. Using an odds ratio of 1.5 (a conservative effect size of 5% difference between groups), 2-sided alpha 0.05, power=90%, and given a ratio of 32:1 (children diagnosed with mc UTI: children with no UTI diagnosis) in the dataset, we require 15,519 children (452 with mcUTI:15,067 without mcUTI) for analysis. Initial examination of the SAIL dataset identified just under 13,000 children less than five years old with UTI between 1999 and 2012. As our sample size calculation requires 452 with UTI, we can be confident of adequate power for this study. However, the true proportion with renal scarring is likely to be less than that reported by Shaikh [8]. Assuming a proportion of 8% with renal scarring in the mcUTI group, and an odds ratio of 1.5 (2.5% difference), 2-sided alpha=0.05, power=90%, and a ratio of 32:1, 35,833 children (1,044 with UTI: 34,789 without UTI) are required for analysis, which is still achievable.

Comparison of systematically versus routinely sampled UTI: This sample size is constrained by the number of children with a systematically sampled microbiologically confirmed UTI by NHS lab (N=374). Using the lower estimate of renal scarring of 8% of routinely sampled UTIs, then a 5% difference (13% in the systematically sampled group) would give 89% power, with a 2-sided alpha=0.05 and a ratio of 32:1 (routinely sampled UTIs 11,968: systematically sampled UTIs 374).

Both analyses use multivariable regression. Using Green's [24] formulae, assuming medium effect sizes (odds ratio=1.5), a 2-sided alpha=0.05, and power=80%, to test 20 predictors in the multivariable regression model we require at least 159 children in total. This suggests that we will be adequately powered for both analyses (given these assumptions) to examine predictors of short and medium-term outcomes.

Statistical analysis

Dataset 1: Routine sampling of UTI

Research question 1: Comparison of medium (up to 5 years) and longer term (≥5 years) outcomes in children with mcUTI versus those without mcUTI in routine sampling.

Baseline variables will be described using appropriate descriptive summaries (N (%), mean (SD), median (interquartile range)) to summarise the population for the main analyses by group 1, 2 and 3 and any marked imbalance between the groups will be identified. There will be no formal testing of between-group differences for any variables at baseline. The main comparative analyses will be carried out at a child level since outcomes relate to an individual's exposure to one or more UTI and will test the null hypothesis that there are no differences in outcomes in the first 5 years and between the age of 5 and 7 years. The primary analysis will consist of a comparison of rate of diagnosis of renal scarring (no renal scarring, renal scarring recorded <5 years, renal scarring recorded 5-7 years) in children

with mcUTI versus children with no mcUTI, using a multinomial regression model. Results will be reported as relative risk ratios alongside 95% confidence intervals (CIs).

We will construct a Direct Acyclic Graph (DAG) to inform our choice of variables to include in the analysis. Confounding variables such as those listed in table 2 and also mcUTI that could be considered to be on the causal pathway will be defined a priori.

We will run multiple mediation analyses using renal scarring as the dependent variable, mcUTI as the mediation variable and confounders as the independent variables. First we will identify the independent variables associated with renal scarring (using an univariable logistic regression and identify the mediation variables (mcUTI or not) that are associated with the significant independent variables. These will all be included in the mediation model. For each of the significant independent variables, two regression models will be performed with and without the mediation variable. We will calculate the indirect effect (and the effect of the mediator) using the logistic regression coefficients from both regression models.

Unadjusted and adjusted relative risk ratios will be estimated, together with 95% CIs. Cox regression will also be performed to model time to first renal scarring diagnosis to allow us to look for this outcome using all available follow-up (at least 7 years). We will estimate hazard ratios with 95% CIs for each exposure group.

Several sensitivity analyses are proposed: The primary outcome will be expanded to include any renal pathology codes due to uncertainty around whether the renal scarring codes are sufficiently sensitive to pick up all cases of renal scarring. In addition we will explore using Read codes from GP notes to pick up additional renal scarring diagnoses. Two effect modifiers were identified as a basis for sub-group analyses for the primary outcome: gender of child and presence of any renal/urological congenital anomalies. These pre-planned analyses will be conducted by the inclusion of appropriate interaction terms in the models.

Secondary outcomes will be analysed using multinomial and time to event models (Table 3). Poisson regression models will be used where the outcome is a count of event (e.g. hospital

admissions, GP consultations, antibiotics prescribed); results will be represented as incidence rate ratios alongside 95% CIs. Additionally we will identify risk factors for renal scarring (as above, plus age at first mcUTI, bacterial type & resistance) in children presenting with a childhood mcUTI. In a sub sample of children linked to GP data, the 14 days prior to the urine sample submission date will be examined to determine whether there was an associated GP consultation. We will then estimate the likely rate of GP associated mcUTIs and the rate of mcUTIs submitted from another setting (such as out of hours, outpatients or as a result of a hospital admission). An identical analysis to the primary outcome will be taken to examine whether the risk of renal scarring differs between those with a mcUTI or not, between different settings and the interaction between the two. Where numbers allow, variation in outcome will be accounted for at the level of the general practice in a multilevel model. We will also describe the levels of urine sampling and incidence of mcUTI from General Practice in the routine data.

Dataset 2: Systematic sampling of UTI

Detailed study data for DUTY and EURICA participants is available including age, gender and deprivation, presenting features, GP diagnosis and acute management. Recruited children are already grouped into mcUTI or no mcUTI through the study microbiology data and outcomes will be compared according to these groups. Five-year in-patient hospital outcomes will be available for all EURICA and DUTY children (the last participant of DUTY study was recruited in April 2012). We will be able to describe serious short-term (30 days and less than 1 year post index consultation) and medium-term (1-5 years) outcomes, including hospital admission, renal imaging, renal scarring, VUR and renal failure outcomes for all children in this group using PEDW data in Wales and NHS Digital hospital data in England. We will be able to describe other outcomes, such as GP consultations, recurrent UTI, dysfunctional voiding, antibiotic prescriptions, hypertension, chronic kidney disease, using GP data on a sub-set. In Wales, Datastore will also be used to look at the urine culture results and organism resistance profile for subsequent UTIs.

We will describe GP diagnosis from study data versus Read codes and acute management from the routine data in GP records for this cohort for later comparisons and also to explore the validity of using routinely collected data in these cases. We will also assess the validity of using Read codes to diagnose UTI against microbiological culture results and agreement will be measured using the Kappa statistic.

Research question 2: Comparison of short- and medium-term outcomes in children with mcUTI: routine versus systematic sampling.

We will compare the outcomes in children with mcUTI identified through routine versus systematic sampling. Children's characteristics, presentation factors, acute management and microbiology results will be described for the groups using appropriate summary statistics. We will compare urine sampling and UTI diagnosis in consultations between routine and systematic sampling. In addition, we will describe blood pressure and creatinine levels for each group if recorded and explore whether comparisons can be made.

Previously mentioned short- and medium-term outcomes will be described by the two groups of routine vs. selective sampling. Predictors of outcome will be examined as before using a multilevel multinomial regression model (no event, 30 days, event 30 days -1 year, event 1-5 years) and again where numbers allowed, variation in outcome will be accounted for at the level of the general practice. Associations between covariates previously described and outcome will firstly be examined. Unadjusted and adjusted relative risk ratios will be estimated, together with 95% CIs. We will compare blood pressure and creatinine levels (where available) across the groups; we expect this data to be limited so will be exploratory.

A detailed statistical analysis plan will be written prior to database lock. The reporting and presentation of results will be in accordance with the [25–27] statements to ensure the comprehensive reporting of our observational non-randomized evaluation of a public health intervention. SPSS and Stata will be used for all analyses [28,29].

Patient and Public Involvement

We have a parent representative (Sarah Jones) who has contributed to all stages of this study. She helped to organise a parent group to discuss information provided to DUTY and EURICA study participants explaining the study and opt-out mechanism. She also provided input on the study website and on the procedure in place to manage contacts made by the participants. During the drafting of the statistical analysis plan we discussed the planned analyses with her, and she identified which of the analyses that she felt would be of most interest to parents of children with suspected UTI. Results will be disseminated via the study website and other channels with the input from our parent representative.

ETHICS AND DISSEMINATION

The governance surrounding dataset one differs from dataset two. Dataset one is an anonymised dataset made available from SAIL databank with only approval required from the IGRP whereas dataset two involves the transfer of identifiable data to data providers which requires ethical approval, section 251 (s251) support, IGRP and IGARD approval. In order to obtain an unbiased sample from the DUTY and EURICA cohort an opt-out consent model is being used, this was supported by both the ethics panel and the CAG panel as justification for this model of consent.

The LUCI study will report the risk of renal scarring for children with and without childhood mcUTI across the whole of Wales. The linking of routinely collected datasets will give us a large cohort including demographic, hospital in-patient and out-patient, GP and microbiology data, allowing us to define mcUTI cases and describe outcomes for all children from both primary and secondary care. Clarifying the link between UTI, renal scarring and long-term complications will inform the management of acutely ill children in primary care, where the need for urine sampling is unclear. Determining the clinical implications of 'missed' cases of UTI through our comparison of children with mcUTI identified through routine and systematic urine sampling will also help to determine the most appropriate urine sampling strategy.

This study maximises the benefits of the previously funded DUTY and EURICA cohorts, representing over 8000 acutely ill children recruited from UK primary care. Significant resources were invested by funders, patients and staff to develop these cohorts. Routine data linkage will allow us to determine longer-term outcomes for these children and to determine risks of adverse outcomes. It will also pave the way for even longer-term follow-up of cohorts of children with UTI (diagnosed both systematically and routinely) which has been identified as a high research priority by NICE.

A lay summary of the results and links to publications will be made available on the University project website. The academic outputs for this study include (i) this protocol paper, (ii) main results from research question one and (ii) main results from research question two. The findings from this study will be of interest to clinicians and policy makers and may influence the management of acutely ill children and childhood UTI.

DECLARATIONS

List of abbreviations

ALF: Anonymised linking field
ALF-E: Anonymised linking field encryption
CIs: Confidence intervals
DAG: Directed acyclic graph
DOB: Date of Birth
ESRF: End-stage renal failure
HES: Hospital Episode Statistics
HRA CAG: Health Research Authority's Confidentiality Advisory Group
HSCIC: Health and Social Care Information Centre
IGARD: Independent Group Advising on the Release of Data
IGRP: Information Governance Review Panel
LSOA: Lower super output area
mcUTI: Microbiological culture urinary tract infection
NICE: National Institute for Health and Clinical Excellence
NIHR HTA: National Institute of Health Research Health Technology Assessment

NISCHR: National Institute for Social Care and Health Research
NWIS: NHS Wales Informatics Service
PEDW: Patient Episode Database for Wales
PRIME: Primary and Emergency Care Research
SAIL: Secure Anonymised Information Linkage
TTP: Trusted third party
UTI: Urinary tract infection
VUR: Vesicoureteric reflux
WDS: Welsh Demographic Service
WECC: Welsh Electronic Cohort of Children

Ethics approval and consent to participate - Ethics approval of the study has been given by the Research Ethics Committee for Wales (16/WA/0166) and the transfer and use of identifiable data has been approved by the Health Research Authority [HRA] Confidentiality Advisory Group [CAG] (16/CAG/0114).

Consent for publication - Not Applicable

Availability of data and material - Not Applicable

Competing Interests - The authors declare that they have no competing interests

Funding - This project has been funded by the Welsh Government through Health and Care Research Wales [Project number 1068].

Authors' contributions- KHu is the chief investigator of the study. All authors have contributed to and are responsible for the final design of the study. FLW is responsible for study management. RCJ and ML are responsible for statistical planning and analysis. LA and HJ are responsible for the data management. All authors have read and approved the final manuscript [KHu, FLW, RCJ, ML, LA, HJ, CB, NF, AH, MH, KH, SP & JV] .

Acknowledgements

We would like to acknowledge the support and input from Sarah Jones, our parent representative for the study. We are also grateful to the DUTY and EURICA participants for their agreement for continued use of their data for this study. The Centre for Trials Research receives funding from Health and Care Research Wales and Cancer Research UK. Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales) receives funding from Health and Care Research Wales. Authors are supported by The Farr Institute CIPHER, funded by Arthritis Research UK, the British Heart Foundation, Cancer Research UK, the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, the Medical Research Council, the National Institute of Health Research, the National Institute for Social Care and Health Research (Welsh Assembly Government), the Chief Scientist Office (Scottish Government Health Directorates), and the Wellcome Trust, (MRC Grant No: MR/K006525/1) and the National Centre for Population Health & Wellbeing Research (NCPHWR).

References

1 Craig JC, Williams GJ, Jones M, *et al.* The accuracy of clinical symptoms and signs for the diagnosis of serious bacterial infection in young febrile children: prospective cohort study of 15 781 febrile illnesses. *BMJ* 2010;**340**:c1594. doi:10.1136/BMJ.C1594

2 Hoberman A, Chao HP, Keller DM, *et al.* Prevalence of urinary tract infection in febrile infants. *J Pediatr* 1993;**123**:17–23.

3 Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am J Med* 2002;**113 Suppl 1A**:5S–13S.

4 Hay AD, Birnie K, Busby J, *et al.* The Diagnosis of Urinary Tract infection in Young children (DUTY): a diagnostic prospective observational study to derive and validate a clinical algorithm for the diagnosis of urinary tract infection in children presenting to primary care with an acute illness. *Health Technol Assess (Rockv)* 2016;**20**. doi:10.3310/hta20510

5 O'Brien K, Edwards A, Hood K, *et al.* Prevalence of urinary tract infection in acutely unwell children in general practice: a prospective study with systematic urine sampling. *Br J Gen Pract* 2013;**63**:e156-64. doi:10.3399/bjgp13X663127

6 Urinary tract infection in under 16s: diagnosis and management | Guidance and guidelines | NICE.

7 Butler CC, O'Brien K, Pickles T, *et al.* Childhood urinary tract infection in primary care: a prospective observational study of prevalence, diagnosis, treatment, and recovery. *Br J Gen Pract* 2015;**65**:e217-23. doi:10.3399/bjgp15X684361

8 Shaikh N, Ewing AL, Bhatnagar S, *et al.* Risk of Renal Scarring in Children With a First Urinary Tract Infection: A Systematic Review. *Pediatrics* 2010;**126**:1084–91. doi:10.1542/peds.2010-0685

9 Coulthard MG, Lambert HJ, Keir MJ. Occurrence of renal scars in children after their first referral for urinary tract infection. *BMJ* 1997;**315**:918–9.

10 Jacobson SH, Eklöf O, Eriksson CG, *et al.* Development of hypertension and uraemia after pyelonephritis in childhood: 27 year follow up. *BMJ* 1989;**299**:703–6.

11 Round J, Fitzgerald AC, Hulme C, *et al.* Urinary tract infections in children and the risk of ESRF. *Acta Paediatr* 2012;**101**:278–82. doi:10.1111/j.1651-2227.2011.02542.x

12 Smellie JM, Prescod NP, Shaw PJ, *et al.* Childhood reflux and urinary infection: a follow-up of 10-41 years in 226 adults. *Pediatr Nephrol* 1998;**12**:727–36.

13 Craig JC, Williams GJ. Denominators do matter: it's a myth--urinary tract infection does not cause chronic kidney disease. *Pediatrics* 2011;**128**:984–5. doi:10.1542/peds.2011-2631

14 Salo J, Ikäheimo R, Tapiainen T, *et al.* Childhood urinary tract infections as a cause of chronic kidney disease. *Pediatrics* 2011;**128**:840–7. doi:10.1542/peds.2010-3520

15 Coulthard MG, Vernon SJ, Lambert HJ, *et al.* A nurse led education and direct access service for the management of urinary tract infections in children: prospective controlled trial. *BMJ* 2003;**327**:656. doi:10.1136/bmj.327.7416.656

16 Jadresic L, Cartwright K, Cowie N, *et al.* Investigation of urinary tract infection in childhood. *BMJ* 1993;**307**:761–4. doi:10.1136/BMJ.307.6907.761

17 van der Voort J, Edwards A, Roberts R, *et al.* The struggle to diagnose UTI in children under two in primary care. *Fam Pract* 1997;**14**:44–8.

18 Hollingworth W, Busby J, Butler CC, *et al.* The Diagnosis of Urinary Tract Infection in Young Children (DUTY) Study Clinical Rule: Economic Evaluation. *Value Heal* 2017;**20**:556–66. doi:10.1016/j.jval.2017.01.003

19 Jones KH, Ford D V., Jones C, *et al.* A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation. *J Biomed Inform* 2014;**50**:196–204. doi:10.1016/j.jbi.2014.01.003

20 Ford D V, Jones KH, Verplancke J-P, *et al.* The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;**9**:157. doi:10.1186/1472-6963-9-157

21 Lyons RA, Jones KH, John G, *et al.* The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009;**9**. doi:10.1186/1472-6947-9-3

22 Robling M, Bekkers M-J, Bell K, *et al.* Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial. *Lancet* 2016;**387**:146–55. doi:10.1016/S0140-6736(15)00392-X

23 Public Health England. UK Standards for Microbiology Investigations. Investigation of Urine. B41. 2018. https://www.gov.uk/government/publications/smi-b-41-investigation-of-urine

24 Green SB. How Many Subjects Does It Take To Do A Regression Analysis. *Multivariate Behav Res* 1991;**26**:499–510. doi:10.1207/s15327906mbr2603_7

25 Gilbert R, Lafferty R, Hagger-Johnson G, *et al.* GUILD: GUidance for Information about Linking Data sets†. *J Public Health (Bangkok)* 2017;:1–8. doi:10.1093/pubmed/fdx037

26 von Elm E, Altman DG, Egger M, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;**335**:806–8. doi:10.1136/bmj.39335.541782.AD

27 Benchimol EI, Smeeth L, Guttmann A, *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med* 2015;**12**:e1001885. doi:10.1371/journal.pmed.1001885

28 IBM. IBM SPSS Statistics for Windows: Version 22.0. 2013.

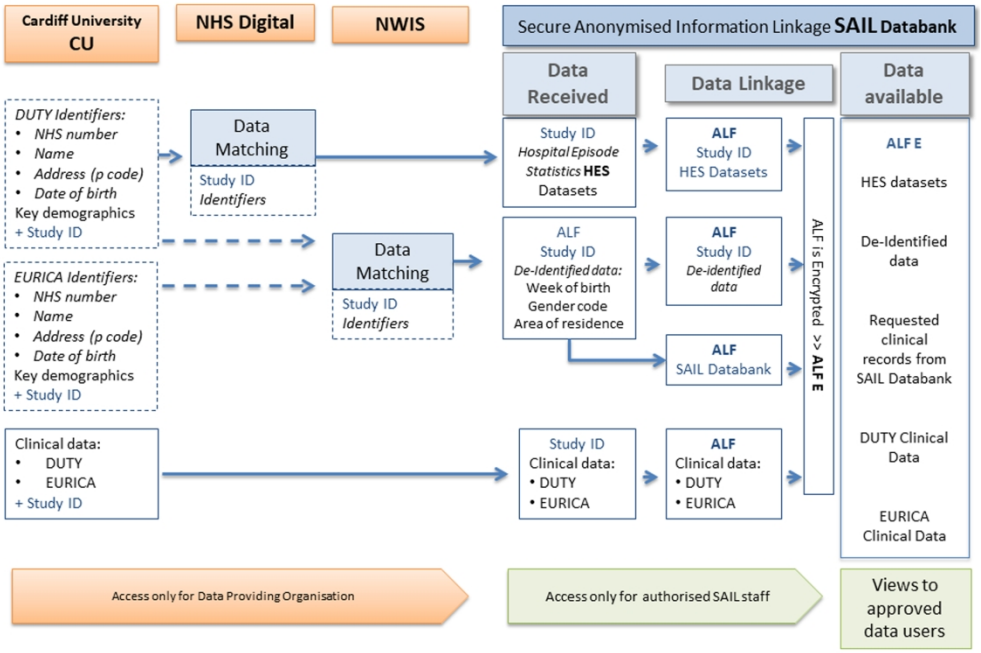
29 StataCorp. Stata Statistical Software: Release 15. 2017. 2017. doi:10.2307/2234838

614

(Title) Figure 1. The data flow for dataset 2.

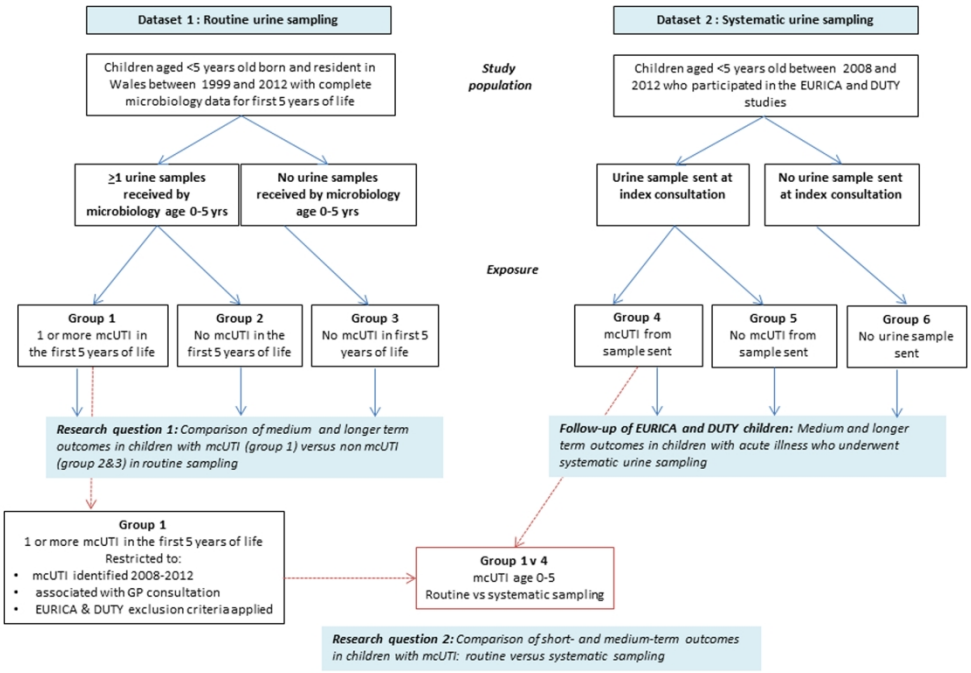
(Legend) ALF- Anonymised Linking Field; NWIS – NHS Wales Informatics Service; HES – Hospital Episode Statistics; SAIL – Secure Anonymised Information Linkage

(Title) Figure 2. Flow chart of study participants



(Title) Figure 1. The data flow for dataset 2.
(Legend) ALF- Anonymised Linking Field; NWIS – NHS Wales Informatics Service; HES – Hospital Episode Statistics; SAIL – Secure Anonymised Information Linkage

297x209mm (300 x 300 DPI)



(Title) Figure 2. Flow chart of study participants

297x209mm (300 x 300 DPI)

BMJ Open

Long-term outcomes of urinary tract infection (UTI) in Childhood (LUCI): protocol for an electronic record-linked cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-024210.R2
Article Type:	Protocol
Date Submitted by the Author:	14-Jan-2019
Complete List of Authors:	Lugg-Widger, Fiona; Cardiff University, Centre for Trials Research Angel, Lianna; Cardiff University, Centre for Trials Research Cannings-John, Rebecca; Cardiff University Centre for Trials Research, Jones, Hywel; Cardiff University, National Centre for Population Health and Wellbeing Research, Division of Population Medicine, Lau, Mandy; Cardiff University, Centre for Trials Research Butler, C; University of Oxford, 3. Nuffield Department of Primary Care Health Sciences Francis, Nick A.; Cardiff University, Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales), Division of Population Medicine, School of Medicine Hay, Alastair; University of Bristol, Centre for Academic Primary Care Heginbothom, Margaret; Public Health Wales Hood, Kerenza; Cardiff University, Centre for Trials Research Paranjothy, Shantini; Cardiff University, National Centre for Population Health and Wellbeing Research, Division of Population Medicine, Vandervoort, Judith; Noah's Arc Childrens Hospital for Wales Hughes, Kathryn; Cardiff University, Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales), Division of Population Medicine, School of Medicine,
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Urology, Paediatrics
Keywords:	Children, Urinary tract infections < UROLOGY, Medical record linkage, Renal scarring, Urine sampling, PRIMARY CARE

Long-term outcomes of urinary tract infection (UTI) in Childhood (LUCI): protocol for an electronic record-linked cohort study

Corresponding author:
Fiona Lugg-Widger LuggFV@cardiff.ac.uk
Address: 4th Floor Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4YS. Centre for Trials Research, College of Biomedical & Life Sciences, Cardiff University, Cardiff, UK.

Lianna Angel[1] AngelL@cardiff.ac.uk
Rebecca Cannings-John[1] CanningsRL@cardiff.ac.uk
Hywel Jones[2] JonesH75@cardiff.ac.uk
Mandy Lau[1] LauTM@cardiff.ac.uk
Christopher C Butler[3] christopher.butler@phc.ox.ac.uk
Nick Francis[4] FrancisNA@cardiff.ac.uk
Alastair D Hay[5] alastair.hay@bristol.ac.uk
Margaret Heginbothom[6] margaret.heginbothom@wales.nhs.uk
Kerenza Hood[1] HoodK1@cardiff.ac.uk
Shantini Paranjothy[2] ParanjothyS@cardiff.ac.uk
Judith Van der Voort[7] judith.vandervoort@wales.nhs.uk
Kathryn Hughes[4] HughesKA6@cardiff.ac.uk

Author Affiliations

1. Centre for Trials Research, College of Biomedical & Life Sciences, Cardiff University, Cardiff, UK.
2. National Centre for Population Health and Wellbeing Research, Division of Population Medicine, Cardiff University
3. Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Primary Care Building, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2 6GG
4. Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales), Division of Population Medicine, School of Medicine, Cardiff University
5. Centre of Academic Primary Care, NIHR School for Primary Care Research, Bristol Medical School: Population Health Sciences, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS
6. Public Health Wales, 2 Capital Quarter, Tyndall Street, Cardiff CF10 4BZ
7. Noah's Arc Childrens Hospital for Wales, Cardiff, CF14 4XW

Word Count: 6810

Keywords: Children; Urinary Tract Infection; Medical record linkage; Renal scarring; Urine sampling; Primary care.

ABSTRACT

Introduction: Current guidelines advise the prompt diagnosis and treatment of UTI in children to improve both short and longer term outcomes. However, the risk of long-term complications following childhood UTI is unclear.

UTI is relatively common but difficult to diagnose in children as symptoms are non-specific. Diagnosis requires a urine sample, but sampling is difficult and infrequent, and it is not clear if sampling should be given greater priority in primary care. The LUCI study will assess the short, medium and longer-term outcomes of childhood UTI associated with routine and systematic sampling practices.

Methods and analysis: Two datasets will be established: The first will consist of routinely collected data (Hospital, GP, Microbiology) from children born and resident in Wales, linked via the Secure Anonymised Information Linkage (SAIL) databank (an 'e-cohort'). Urine sampling in this dataset reflects normal practice 'routine sampling'. Outcomes (including renal scarring, hypertension, end-stage renal failure(ESRF), hospital admissions, GP consultations, antibiotic prescriptions) for children with at least one UTI confirmed with microbiological culture (mcUTI) or no mcUTI before the age of 5 will be compared.

The second will combine data from two prospective observational studies('DUTY' & 'EURICA') employing systematic urine sampling for children presenting to primary care with acute, undifferentiated illness, linked to routine data via SAIL (Wales) and NHS Digital (England). Outcomes (as above, plus features of mcUTI) for children with a mcUTI in this dataset, identified through systematic urine sampling, will be compared to those with a mcUTI identified through routine urine sampling (dataset 1).

Ethics and dissemination: The study protocol has been approved by NHS Wales Research Ethics Committee and the Health Research Authority's Confidentiality Advisory Group. Methods of innovative study design and findings will be disseminated through peer-review

journals and conferences. Results will be of interest to clinical and policy stakeholders in the UK.

Strengths and limitations of this study:

- Use of routinely collected data in the study allows the identification of rare chronic outcomes, from large numbers of children at risk.
- This multi-sourced dataset will allow a comparison of outcomes over 5 years for children with and without microbiologically confirmed UTI (mcUTI) according to routine clinical practice; and compare outcomes in these groups with those observed in high quality research data using systematic urine sampling.
- This study will help to prioritise interventions to improve early diagnosis, sampling and treatment, potentially improving health outcomes and reducing NHS costs.
- Using routinely collected data relies on the quality of coding and availability of data.
- Using routinely collected data limits the information available on the children and their outcomes.

INTRODUCTION

Urinary tract infections (UTI) are a common cause of acute illness in children and an important contributor to hospital admissions for serious bacterial infection. [1–7] In UK primary care, UTI is the cause of approximately 6% of acute illness consultations in children less than five years old. [5,7]

Childhood UTI can lead to renal scarring. [6,8,9] Renal scarring has been linked with long-term complications including hypertension, pre-eclampsia and renal failure. [10–12] Clinical guidelines promote the prompt diagnosis and treatment of UTI as a measure to prevent renal scarring and longer-term complications. [6] It is not clear what the risk of longer-term complications are for children with UTI. A systematic review in 2010 found that the prevalence

of renal scarring following first childhood UTI was 15%. [8] Most included studies were conducted in secondary care and most required fever for inclusion in the study [8]. These are likely to represent more serious UTIs than those presenting in primary care, and this rate of renal scarring may not apply to all children presenting with UTI. The risk of long-term complications (such as hypertension and renal failure) as a result of renal scarring following childhood UTI is also not clear, with some researchers questioning this association. [6,13,14] A review, commissioned by the National Institute for Health and Clinical Excellence (NICE), concluded that 'there are no appropriate studies that accurately estimate the risks of long term complications as a result of childhood UTI'. [6]

A urine sample is necessary to confirm the presence of UTI in childhood as the presenting symptoms are non-specific and similar to those found in many common childhood illnesses (e.g. respiratory infections, gastroenteritis, tonsillitis, ear infections) such as high temperature, poor feeding, vomiting, lethargy and irritability. [6] Furthermore, microbiological confirmation is important as some children with UTI require invasive investigations. [6] Children who become seriously ill and who are assessed in emergency departments or admitted to hospital will usually have their urine sampled. [1] However, in primary care, where most acute illness is seen, urine is infrequently sampled. It is estimated that urine is sampled in fewer than 2% of acute illness consultations with children under five years old in the UK. [4] Studies have suggested that many cases of childhood UTI are missed in primary care. [15–17] Routine practice (urine sampling based on clinician suspicion) is likely to miss more than two thirds of UTIs in primary care. [5,7] The clinical implications of missed childhood UTIs are not known. Increasing urine sampling in primary care is likely to increase the diagnosis of childhood mcUTI, and is advocated by current guidelines, but it is not clear whether this is an appropriate strategy, whether outcomes for children would improve or to what extent it should be increased. [4–6,15,18]

Clarification is needed on two issues: First, to determine what the longer-term outcomes are following childhood UTI (including those identified in primary care), and second, to determine whether outcomes vary according to sampling practice.

In this study we will describe clinical outcomes for all children with one or more mcUTI aged less than five years old, compared to those with no mcUTI, using NHS laboratory data from across Wales. We will examine the risk factors for being diagnosed with renal scarring following mcUTI.

We will also describe longer-term follow up of clinical outcomes for at least five years following participation in two UK prospective cohort studies of acutely ill children with systematic urine sampling in primary care, the DUTY and EURICA studies. [5,7] We will compare the outcomes of those with mcUTI identified through these studies (systematic urine sampling) with those identified through routine practice.

METHODS AND DESIGN

Research objectives

The LUCI Study will use data linkage of routinely collected datasets and data from two cohorts of participants to answer two main research questions:

Research Question 1: Through routine sampling, do children who have experienced a mcUTI aged <5 years old have worse outcomes (medium (0-5 years); longer-term (>5 years)) compared to children who have not experienced a mcUTI?

Research Question 2: Are short-term (<12 months) and medium-term (1-4 years) outcomes different for children with childhood mcUTI identified through systematic sampling compared to routine sampling (standard, clinician-led sampling)?

Study Design

This is a data linkage study comprising two overarching datasets of children. Dataset 1 will comprise routinely collected health data from children born and resident in Wales. Children in

this dataset will have had urine sampled according to routine practice. Routine data will be available on all children for seven years, and longer for some (i.e. children will be followed up until the date of data abstraction). Dataset 1 will be used to answer Research Question 1.

Dataset 2 will be children who participated in the EURICA or DUTY studies. Children in this dataset had their urine systematically sampled (all children presenting with an acute illness were asked to provide a urine sample). DUTY and EURICA children will be followed-up by linking records to routinely collected health data from England (using NHS Digital) and Wales (using SAIL). Follow-up will be for a minimum of five years, and longer where data is available. Dataset 2 will be used to describe longer term follow-up for DUTY and EURICA study children. Those with mcUTI in Dataset 2 will be compared to those with mcUTI identified in Dataset 1 to answer Research Question 2.

The study formally started in October 2016 and will report to funder in June 2019. A summary of the data sources is provided in Table 1.

Table 1: Sources of Data

Data Provider	Data source	Dates available from - to	Indicative / key data items	Dataset	
				Dataset 1 (routine sampling)	Dataset 2 (systematic sampling)
SAIL (Wales)	GP	Jan 1994 – Oct 2016	Secondary outcomes including antibiotic prescriptions, GP consultations, chronic kidney disease, hypertension	✓	✓
	Patient Episode Database for Wales (PEDW)	Jan 1994 – April 2017	Primary & secondary outcomes & covariates including renal scarring, hospital admission, end-stage renal failure, VUR, renal/bladder surgery	✓	✓
	Welsh Demographic Service (WDS)	Jan 1994 - April 2017	Demographics	✓	✓
	Welsh Electronic Cohort of Children (WECC)	Jan 1994 - Sept 2011	Defines a child as “born in Wales”	✓	✓
	Datastore (data repository storing microbiology culture data from the Laboratory Information Management Systems in Wales)	2005 - 2014	Defines a microbiologically confirmed UTI	✓	✓
	Outpatient data	Jan 1994 - April 2017	Primary outcome – renal scarring	✓	✓
Individual Health Boards	Radiology Data	Jan 1994 – Jan 2017	Validation of the primary outcome (renal scarring)	✓	
NHS Digital (England)	Admitted Patient Care	April 2008 - Mar 2017	Primary & secondary outcomes & covariates including renal scarring, hospital admission, end-stage renal failure, VUR renal/bladder surgery		✓
	Outpatient	April 2008 - Mar 2017	Primary outcome – renal scarring		✓

Bristol University (England and Wales)	DUTY	April 2010 – April 2012	Baseline characteristics, UTI status, presenting symptoms & signs, initial clinical management		✓
Cardiff University (Wales)	EURICA	July 2008 – July 2010	Baseline characteristics, UTI status, presenting symptoms & signs, initial clinical management		✓

162

Data providers and datasets

The EURICA and DUTY Studies

This work builds on two large cohort studies of acutely ill children, aged less than five years old, presenting in primary care, in which mcUTI status was determined using systematic urine sampling [5,7]. In both studies, clinical and demographic data were collected and urine samples requested from all children included in the study and analysed in NHS microbiology laboratories. In the EURICA study, 1003 children were recruited from GP practices in Wales between March 2008 and July 2010. 5.9% children had mcUTI. Participants had a GP notes review at 6 months. In the DUTY study, 7163 children were recruited from GP practices in England and Wales between April 2010 and April 2012 and 5.6% children had mcUTI confirmed in NHS labs. An algorithm to predict UTI based on presenting symptoms and signs was developed. Neither study had sufficient follow-up to determine whether renal investigations to look for renal scarring had been undertaken or found. EURICA was funded by NISCHR (now Health and Care Research Wales) and sponsored by Cardiff University and DUTY was NIHR HTA funded and sponsored by Bristol University.

SAIL Databank

The study will use the Secure Anonymised Information Linkage (SAIL) Databank to access routinely collected data outlined in Table 1. SAIL is a repository for a broad range of routinely collected health and population data in Wales. SAIL will also act as a data safe haven for the clinical DUTY and EURICA datasets and data made available from NHS Digital and Individual Health Boards. All data will be accessed via the SAIL Gateway following Information Governance Review Panel (IGRP) approval. SAIL Databank does not handle any identifiable data therefore all data will be anonymised including data transferred from other information centres [19–21].

NHS Digital

NHS Digital is the trading name of the Health and Social Care Information Centre (HSCIC). This study will access Hospital Episode Statistics (HES) data for participants of the DUTY and EURICA study. All available Inpatient and Outpatient records belonging to each study participant will be requested and approved by the Independent Group Advising on the Release of Data (IGARD) Panel. The data requested includes diagnosis, procedures and length of episode according to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems [ICD-10] codes.

Public Health Wales

Public Health Wales will provide a data extract of urine microbiology culture results from all microbiology laboratories in Wales (Datastore) for use with this project. This will be transferred to SAIL.

Individual Health Boards

Health boards in Wales will be approached to access anonymised radiology data for patients in dataset 1. A one off data extract of patient-level attendance data for patients born between 01/01/1994 and 31/12/2012 that attended radiology between 1994 and 2016 will be transferred to SAIL. Data extracted includes examination performed, attendance data and the radiology report.

Opportunity to opt-out (dataset two)

Dataset one uses routinely collected data that is fully anonymised so we do not require individual consent in order to access these data. Dataset two involves participants from the DUTY and EURICA study and therefore requires section 251 (s251) support of the 2006 NHS

Act approval from the Health Research Authority's Confidentiality Advisory Group (HRA CAG) to pass identifiable participant data legally held by Cardiff and Bristol University (as sponsors of the studies) to the data providers. This is using an opt-out/dissent model instead of obtaining further consent. In order to provide the opportunity for participants to dissent a letter has been developed to be sent to all parents of participants. Participants have the opportunity to contact the study team through an online web-form, by email, text or telephone (details provided on website) and register their dissent. These participants who register their dissent will not be included in any of the datasets sent to the information centres and therefore will not appear in the datasets for analysis. A participant representative was consulted on the layout, wording and level of information contained in the participant opt-out letter and on the study website. A key consideration was to communicate the data transfer process. The final letter was approved by both an NHS Research Ethics Committee and CAG committee as part of overall governance approval for the study.

Data Matching

For included data sets, data will be sent to NHS Digital and SAIL (via NWIS their trusted third party (TTP)) for matching using a combination of NHS number, name, address and date of birth. Matching with NHS Digital data will be by exact matching on NHS Number, Date of Birth and Postcode. This has been conducted in other studies and achieved a high match rate (99.6%) [22]. NWIS match using NHS Number, Date of Birth, Sex, Postcode. Data matching by NHS Digital will be separate to matching conducted by NWIS on behalf of SAIL. We would expect only a small number of participants matched to both English and Welsh NHS records however there is the possibility of this for those using services across the border to their current address.

The anonymised dataset

For DUTY and EURICA participants NHS Digital will retain the unique study ID assigned by the DUTY/EURICA study, with the HES records after the matching process before it is sent to SAIL Databank. This ID will be retained for data transfer to ensure the data can be linked. The same applies to data sent to NWIS. Requested data will then be linked for each participant via SAIL using the unique study ID. The DUTY and EURICA clinical datasets will be transferred to SAIL following a process of de-identification. The data flow is shown in Figure 1. The study ID will be replaced by an anonymised linking field (ALF) to enable all datasets to be linked at an individual level. The resulting dataset will contain clinical variables (from DUTY, EURICA, SAIL and NHS Digital), de-identified demographic variables (e.g. week of birth and lower super output area - LSOA) and the ALF. The key between study ID and ALF will be retained and encrypted as a further safeguard. The ALF is further encrypted (ALF-E) so that datasets and records cannot be linked across multiple projects a researcher has access to.

Study Participants

A flow chart of participants in the study is shown in Figure 2. Dataset 1 (routine sampling) will be children born and resident in Wales for the first 5 years of life; who were less than 5 years old between 1999 and 2012. The main analysis will be on children born between 01/01/2005 and 31/12/2009 to ensure that all of the first five years of life are covered by the dates when Datastore is available.

Dataset 2 (systematic sampling from DUTY and EURICA) will be children who participated in the EURICA or DUTY studies, who were not withdrawn from the study and when provided with the opportunity to opt-out, remained in the study. For research question 2 when comparing children with a mcUTI from Dataset 1 and 2, children in Group 1 will be selected to match the DUTY and EURICA study eligibility criteria as closely as possible within the constraints of using routine data. Datasets 1 and 2 will be limited to children with a mcUTI associated with a GP consultation between March 2008 and April 2012. For both datasets, the child's mcUTI, associated with a GP consultation (defined as within 14 days prior to the sample) will be

identified and defined as the index consultation. To limit the potential transfer of GP systematic sampling behaviour, children from Group 1 with index consultations between 2008 and 2012 at practices which participated in the EURICA study will be flagged as will children with index consultations between 2010 and 2012 at practices which participated in the DUTY study. Children will only be included once in each study period (i.e. a child with a sample sent within the EURICA study period could also appear in the DUTY study period). In addition, we will apply the DUTY study exclusion criteria (where possible) to Dataset 1 to make them directly comparable with Dataset 2; therefore children will be excluded from the routine sampling cohort if any criterion met:

- Known neurogenic bladder (e.g. spina bifida) or previous bladder surgery
- Prescribed antibiotics in the 7 days prior to presentation
- Taking immunosuppressant medication
- Using urinary catheters

Children with a prior history of UTI or vesicoureteric reflux (VUR) will not be excluded (and were not excluded from the DUTY or EURICA studies) and we will explore the impact of these risk factors on outcomes.

Exposure

UTI cases will be based on NHS laboratory results. For Dataset 1, this will be through microbiological culture data downloaded from Datastore. These data represent samples (from both community and hospital settings) which have been classified as positive or negative by NHS laboratories according to their standard operating procedures. We do not know how urine was sampled, and this is likely to vary between settings. In most cases, these are likely to be clean catch samples, but may include urine collection pads or bags (particularly in community

samples) as recommended by NICE; or catheter or suprapubic aspiration (SPA) in hospital samples.[6] NHS laboratories take into consideration the nature of the urine sample in their reporting.[23] For Dataset 2, we will use the results of microbiological culture from NHS laboratories collected during the DUTY and EURICA studies as some participants were from England (Datastore is Wales only). For Dataset 2, the presence of significant bacteriuria (pure or predominant growth of 100,000 cfu/ml of urine) will be used to define UTI.

For Dataset 1 we define the exposure period as <5 years and will be grouped as follows: (Figure 1).

Group 1: children with at least one mcUTI before their 5th birthday or before outcome of interest

Group 2: children with at least one urine sample but no mcUTI before their 5th birthday or before outcome of interest

Group 3: children with no urine samples before their 5th birthday or before outcome of interest

Exposure is a discrete time-varying covariate (0 (group 2 or 3 no UTI or no sample respectively) until first exposure of group 1 (mcUTI) thereafter) and each patient's exposure status will be taken at the point of each outcome; otherwise the exposure status of the child at their 5th birthday will be taken.

For the main analyses, Groups 2 and 3 will be considered together as having no microbiologically confirmed UTI. Dataset 2 will similarly be divided into three groups based on their index consultation when recruited into the DUTY and EURICA studies:

Group 4: children with a mcUTI

Group 5: children who had a urine sample but no mcUTI

Group 6: children who had no urine sample

Study variables

Table 2 shows a breakdown of the baseline data and possible covariates available for children and maternal characteristics from the data collection forms for EURICA and DUTY and WDS, WECC, and for a subset with GP records. The study outcomes are summarised in Table 3.

Table 2. Child and maternal characteristics

Risk factor	**Dataset 1: Routine sampling: Source**	**Dataset 2: Systematic sampling: Source**
Age of child at first urine sample (group 1/2 only) or index consultation (Dataset 2)	Datastore	EURICA & DUTY study data
Gender	WECC	EURICA & DUTY study data
Ethnicity	WECC	WECC & DUTY study data
Deprivation quintile at birth (taken from postcode at birth)	WDS: WIMD & Townsend score	Townsend score; EURICA & DUTY study data
Maternal age at birth (years) (category)	WECC	EURICA & DUTY Welsh participants only: WECC
Birth weight (grams) (category)	WECC	EURICA & DUTY Welsh participants only: WECC
Gestational age at birth (weeks) (category)	WECC	EURICA & DUTY Welsh participants only: WECC
Ever breastfed	WECC	EURICA & DUTY study data
Uropathogen (Enterobacteriaceae or other) (mcUTI only)	Datastore	EURICA & DUTY study data
Antimicrobial resistance (mcUTI only)	Datastore	EURICA & DUTY study data
Congenital malformations (con mals)		
known to be associated with UTI/renal scarring:	PEDW	
i. Spina bifida/neuro bladder		N/A: DUTY exclusion
ii. Renal/urinary system con mals including vesico-uretero-renal reflux (VUR)		EURICA & DUTY study data
possibly associated with UTI/renal scarring:		
i. Downs Syndrome	PEDW; WECC	EURICA & DUTY study data
ii. Cerebral palsy/other paralytic syndromes	WECC	EURICA & DUTY study data
Congenital malformations - Major/Minor	WECC	EURICA & DUTY Welsh participants only: WECC

Risk factor	Dataset 1: Routine sampling: Source	Dataset 2: Systematic sampling: Source
Comorbidities		
i. Diabetes diagnosed under the age of 5 years	PEDW	EURICA & DUTY study data*
ii. Renal or urogenital surgery	PEDW	EURICA study data* DUTY exclusion
iii. Cancer	PEDW	EURICA & DUTY study data*
iv. Immunosuppressive disease	PEDW	EURICA & DUTY study data*
v. Circumcision (aged <5 years)	PEDW; GP	EURICA & DUTY study data*
Factors for follow-up of study participants & Research Question 2		
Symptoms & signs at index consultation	-	EURICA & DUTY study data
Management at index consultation	GP	EURICA & DUTY study data
Antenatal ultrasound urinary system abnormalities	-	EURICA & DUTY study data
Family history of UTI/urinary system problems	-	EURICA & DUTY study data
Recent antibiotics (7 days prior to index consultation)	GP	EURICA & DUTY* study data

* at time of index consultation

Table 3. Study Outcomes

	Data source		
	PEDW (All Wales)	**GP**	**Datastore (All Wales)**
Primary outcome			
Renal scarring	✓		
Sensitivity analyses			
Any renal pathology codes	✓		
GP renal scarring codes		✓	
Secondary outcomes			
Hospital admissions	✓		
Day cases	✓		
Renal/urological surgery	✓		
Hypertension	✓	✓	
Chronic kidney disease	✓	✓	
Renal failure	✓		
UTIs	✓	✓	
Renal imaging		✓	
GP consultations		✓	
Antibiotics		✓	
Dysfunctional voiding		✓	
Microbiologically confirmed UTI (5-7yrs follow up)			✓

Follow-up

Follow-up of all children in Datasets 1 and 2 will continue until the first occurrence of any of: outcome, migration, death or end of follow-up; and for the sub-analysis of GP data, if the

patient leaves the GP practice linked to SAIL or the last data collection from the general practice. For the analysis of Research Question 1, using only children whose whole first five years of life were covered by the dates that Datastore was available (excluding children whos first five years of life fall outside of Datastore availability) and limiting DOBs to 1/1/2005 to 31/12/2009, will mean that shortest follow-up period will be 7 years. Follow-up will be longer where data is available. For Research Question 2, we will examine outcomes at 30 days and at 1 year (short term), 1-5 years (medium-term) and >5 years (long-term) after the index consultation.

Analysis

Sample size

Comparison of renal scarring in children routinely sampled for UTI with and without mcUTI: The sample size is based on the outcome of renal scarring of children with and without mcUTI and taken as 15% [8]. Using an odds ratio of 1.5 (a conservative effect size of 5% difference between groups), 2-sided alpha 0.05, power=90%, and given a ratio of 32:1 (children diagnosed with mc UTI: children with no UTI diagnosis) in the dataset, we require 15,519 children (452 with mcUTI:15,067 without mcUTI) for analysis. Initial examination of the SAIL dataset identified just under 13,000 children less than five years old with UTI between 1999 and 2012. As our sample size calculation requires 452 with UTI, we can be confident of adequate power for this study. However, the true proportion with renal scarring is likely to be less than that reported by Shaikh [8]. Assuming a proportion of 8% with renal scarring in the mcUTI group, and an odds ratio of 1.5 (2.5% difference), 2-sided alpha=0.05, power=90%, and a ratio of 32:1, 35,833 children (1,044 with UTI: 34,789 without UTI) are required for analysis, which is still achievable.

Comparison of systematically versus routinely sampled UTI: This sample size is constrained by the number of children with a systematically sampled microbiologically confirmed UTI by NHS lab (N=374). Using the lower estimate of renal scarring of 8% of routinely sampled UTIs,

then a 5% difference (13% in the systematically sampled group) would give 89% power, with a 2-sided alpha=0.05 and a ratio of 32:1 (routinely sampled UTIs 11,968: systematically sampled UTIs 374).

Both analyses use multivariable regression. Using Green's [24] formulae, assuming medium effect sizes (odds ratio=1.5), a 2-sided alpha=0.05, and power=80%, to test 20 predictors in the multivariable regression model we require at least 159 children in total. This suggests that we will be adequately powered for both analyses (given these assumptions) to examine predictors of short and medium-term outcomes.

Statistical analysis

Dataset 1: Routine sampling of UTI

Research question 1: Comparison of medium (up to 5 years) and longer term (≥5 years) outcomes in children with mcUTI versus those without mcUTI in routine sampling.

Baseline variables will be described using appropriate descriptive summaries (N (%), mean (SD), median (interquartile range)) to summarise the population for the main analyses by group 1, 2 and 3 and any marked imbalance between the groups will be identified. There will be no formal testing of between-group differences for any variables at baseline. The main comparative analyses will be carried out at a child level since outcomes relate to an individual's exposure to one or more UTI and will test the null hypothesis that there are no differences in outcomes in the first 5 years and between the age of 5 and 7 years. The primary analysis will consist of a comparison of rate of diagnosis of renal scarring (no renal scarring, renal scarring recorded <5 years, renal scarring recorded 5-7 years) in children with mcUTI versus children with no mcUTI, using a multinomial regression model. Results will be reported as relative risk ratios alongside 95% confidence intervals (CIs). A survival model will also be performed to model time to first renal scarring diagnosis taking into account competing risks (such as deaths and migration) and differences in time-at-risk and to allow us to look for this outcome using all

available follow-up for each child (at least 7 years). We will estimate hazard ratios with 95% CIs for each exposure group.

We will construct a Direct Acyclic Graph (DAG) to inform our choice of variables to include in the analysis. Confounding variables such as those listed in table 2 and also mcUTI that could be considered to be on the causal pathway will be defined a priori. We will run multiple mediation analyses using renal scarring as the dependent variable, mcUTI as the mediation variable and confounders as the independent variables. First we will identify the independent variables associated with renal scarring (using an univariable logistic (where scarring is rare) or log-linear regression model (where scarring is common)) and identify the mediation variables (mcUTI or not) that are associated with the significant independent variables. These will all be included in the mediation model. For each of the significant independent variables, two regression models will be performed with and without the mediation variable. We will calculate the indirect effect (and the effect of the mediator) using the regression coefficients from both regression models.

Unadjusted and adjusted relative risk ratios will be estimated, together with 95% CIs.

Several sensitivity analyses are proposed: The primary outcome will be expanded to include any renal pathology codes due to uncertainty around whether the renal scarring codes are sufficiently sensitive to pick up all cases of renal scarring. In addition we will explore using Read codes from GP notes to pick up additional renal scarring diagnoses. Two effect modifiers were identified as a basis for sub-group analyses for the primary outcome: gender of child and presence of any renal/urological congenital anomalies. These pre-planned analyses will be conducted by the inclusion of appropriate interaction terms in the models.

Secondary outcomes will be analysed using multinomial and time to event models (Table 3). Poisson regression models will be used where the outcome is a count of event (e.g. hospital admissions, GP consultations, antibiotics prescribed); results will be represented as incidence

rate ratios alongside 95% CIs. Additionally we will identify risk factors for renal scarring (as above, plus age at first mcUTI, bacterial type & resistance) in children presenting with a childhood mcUTI. In a sub sample of children linked to GP data, the 14 days prior to the urine sample submission date will be examined to determine whether there was an associated GP consultation. We will then estimate the likely rate of GP associated mcUTIs and the rate of mcUTIs submitted from another setting (such as out of hours, outpatients or as a result of a hospital admission). An identical analysis to the primary outcome will be taken to examine whether the risk of renal scarring differs between those with a mcUTI or not, between different settings and the interaction between the two. Where numbers allow, variation in outcome will be accounted for at the level of the general practice in a multilevel model. We will also describe the levels of urine sampling and incidence of mcUTI from General Practice in the routine data.

Dataset 2: Systematic sampling of UTI

Detailed study data for DUTY and EURICA participants is available including age, gender and deprivation, presenting features, GP diagnosis and acute management. Recruited children are already grouped into mcUTI or no mcUTI through the study microbiology data and outcomes will be compared according to these groups. Five-year in-patient hospital outcomes will be available for all EURICA and DUTY children (the last participant of DUTY study was recruited in April 2012). We will be able to describe serious short-term (30 days and less than 1 year post index consultation) and medium-term (1-5 years) outcomes, including hospital admission, renal imaging, renal scarring, VUR and renal failure outcomes for all children in this group using PEDW data in Wales and NHS Digital hospital data in England. We will be able to describe other outcomes, such as GP consultations, recurrent UTI, dysfunctional voiding, antibiotic prescriptions, hypertension, chronic kidney disease, using GP data on a sub-set. In Wales, Datastore will also be used to look at the urine culture results and organism resistance profile for subsequent UTIs.

We will describe GP diagnosis from study data versus Read codes and acute management from the routine data in GP records for this cohort for later comparisons and also to explore

the validity of using routinely collected data in these cases. We will also assess the validity of using Read codes to diagnose UTI against microbiological culture results and agreement will be measured using the Kappa statistic.

Research question 2: Comparison of short- and medium-term outcomes in children with mcUTI: routine versus systematic sampling.

We will compare the outcomes in children with mcUTI identified through routine versus systematic sampling. Children's characteristics, presentation factors, acute management and microbiology results will be described for the groups using appropriate summary statistics. We will compare urine sampling and UTI diagnosis in consultations between routine and systematic sampling. In addition, we will describe blood pressure and creatinine levels for each group if recorded and explore whether comparisons can be made.

Previously mentioned short- and medium-term outcomes will be described by the two groups of routine vs. selective sampling. Predictors of outcome will be examined as before using a multilevel multinomial regression model (no event, 30 days, event 30 days -1 year, event 1-5 years) and again where numbers allowed, variation in outcome will be accounted for at the level of the general practice. Associations between covariates previously described and outcome will firstly be examined. Unadjusted and adjusted relative risk ratios will be estimated, together with 95% CIs. We will compare blood pressure and creatinine levels (where available) across the groups; we expect this data to be limited so will be exploratory.

A detailed statistical analysis plan will be written prior to database lock. The reporting and presentation of results will be in accordance with the [25–27] statements to ensure the comprehensive reporting of our observational non-randomized evaluation of a public health intervention. SPSS and Stata will be used for all analyses [28,29].

Patient and Public Involvement

We have a parent representative (Sarah Jones) who has contributed to all stages of this study. She helped to organise a parent group to discuss information provided to DUTY and EURICA study participants explaining the study and opt-out mechanism. She also provided input on the study website and on the procedure in place to manage contacts made by the participants. During the drafting of the statistical analysis plan we discussed the planned analyses with her, and she identified which of the analyses that she felt would be of most interest to parents of children with suspected UTI. Results will be disseminated via the study website and other channels with the input from our parent representative.

ETHICS AND DISSEMINATION

The governance surrounding dataset one differs from dataset two. Dataset one is an anonymised dataset made available from SAIL databank with only approval required from the IGRP whereas dataset two involves the transfer of identifiable data to data providers which requires ethical approval, section 251 (s251) support, IGRP and IGARD approval. In order to obtain an unbiased sample from the DUTY and EURICA cohort an opt-out consent model is being used, this was supported by both the ethics panel and the CAG panel as justification for this model of consent.

The LUCI study will report the risk of renal scarring for children with and without childhood mcUTI across the whole of Wales. The linking of routinely collected datasets will give us a large cohort including demographic, hospital in-patient and out-patient, GP and microbiology data, allowing us to define mcUTI cases and describe outcomes for all children from both primary and secondary care. Clarifying the link between UTI, renal scarring and long-term complications will inform the management of acutely ill children in primary care, where the need for urine sampling is unclear. Determining the clinical implications of 'missed' cases of UTI through our comparison of children with mcUTI identified through routine and systematic urine sampling will also help to determine the most appropriate urine sampling strategy. This study maximises the benefits of the previously funded DUTY and EURICA cohorts, representing over 8000 acutely ill children recruited from UK primary care. Significant

resources were invested by funders, patients and staff to develop these cohorts. Routine data linkage will allow us to determine longer-term outcomes for these children and to determine risks of adverse outcomes. It will also pave the way for even longer-term follow-up of cohorts of children with UTI (diagnosed both systematically and routinely) which has been identified as a high research priority by NICE.

A lay summary of the results and links to publications will be made available on the University project website. The academic outputs for this study include (i) this protocol paper, (ii) main results from research question one and (ii) main results from research question two. The findings from this study will be of interest to clinicians and policy makers and may influence the management of acutely ill children and childhood UTI.

DECLARATIONS

List of abbreviations

ALF: Anonymised linking field

ALF-E: Anonymised linking field encryption

CIs: Confidence intervals

DAG: Directed acyclic graph

DOB: Date of Birth

ESRF: End-stage renal failure

HES: Hospital Episode Statistics

HRA CAG: Health Research Authority's Confidentiality Advisory Group

HSCIC: Health and Social Care Information Centre

IGARD: Independent Group Advising on the Release of Data

IGRP: Information Governance Review Panel

LSOA: Lower super output area

mcUTI: Microbiological culture urinary tract infection

NICE: National Institute for Health and Clinical Excellence

NIHR HTA: National Institute of Health Research Health Technology Assessment

NISCHR: National Institute for Social Care and Health Research

NWIS: NHS Wales Informatics Service

PEDW: Patient Episode Database for Wales

PRIME: Primary and Emergency Care Research

SAIL: Secure Anonymised Information Linkage

TTP: Trusted third party

UTI: Urinary tract infection

VUR: Vesicoureteric reflux

WDS: Welsh Demographic Service

WECC: Welsh Electronic Cohort of Children

Ethics approval and consent to participate - Ethics approval of the study has been given by the Research Ethics Committee for Wales (16/WA/0166) and the transfer and use of identifiable data has been approved by the Health Research Authority [HRA] Confidentiality Advisory Group [CAG] (16/CAG/0114).

Consent for publication - Not Applicable

Availability of data and material - Not Applicable

Competing Interests - The authors declare that they have no competing interests

Funding - This project has been funded by the Welsh Government through Health and Care Research Wales [Project number 1068].

Authors' contributions- KHu is the chief investigator of the study. All authors have contributed to and are responsible for the final design of the study. FLW is responsible for study management. RCJ and ML are responsible for statistical planning and analysis. LA and HJ are responsible for the data management. All authors have read and approved the final manuscript [KHu, FLW, RCJ, ML, LA, HJ, CB, NF, AH, MH, KH, SP & JV] .

Acknowledgements

We would like to acknowledge the support and input from Sarah Jones, our parent representative for the study. We are also grateful to the DUTY and EURICA participants for their agreement for continued use of their data for this study. The Centre for Trials Research

receives funding from Health and Care Research Wales and Cancer Research UK. Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales) receives funding from Health and Care Research Wales. Authors are supported by The Farr Institute CIPHER, funded by Arthritis Research UK, the British Heart Foundation, Cancer Research UK, the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, the Medical Research Council, the National Institute of Health Research, the National Institute for Social Care and Health Research (Welsh Assembly Government), the Chief Scientist Office (Scottish Government Health Directorates), and the Wellcome Trust, (MRC Grant No: MR/K006525/1) and the National Centre for Population Health & Wellbeing Research (NCPHWR).

References

1 Craig JC, Williams GJ, Jones M, *et al.* The accuracy of clinical symptoms and signs for the diagnosis of serious bacterial infection in young febrile children: prospective cohort study of 15 781 febrile illnesses. *BMJ* 2010;**340**:c1594. doi:10.1136/BMJ.C1594

2 Hoberman A, Chao HP, Keller DM, *et al.* Prevalence of urinary tract infection in febrile infants. *J Pediatr* 1993;**123**:17–23.

3 Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am J Med* 2002;**113 Suppl 1A**:5S–13S.

4 Hay AD, Birnie K, Busby J, *et al.* The Diagnosis of Urinary Tract infection in Young children (DUTY): a diagnostic prospective observational study to derive and validate a clinical algorithm for the diagnosis of urinary tract infection in children presenting to primary care with an acute illness. *Health Technol Assess (Rockv)* 2016;**20**. doi:10.3310/hta20510

5 O'Brien K, Edwards A, Hood K, *et al.* Prevalence of urinary tract infection in acutely unwell children in general practice: a prospective study with systematic urine sampling. *Br J Gen Pract* 2013;**63**:e156-64. doi:10.3399/bjgp13X663127

6 Urinary tract infection in under 16s: diagnosis and management | Guidance and guidelines | NICE.

7 Butler CC, O'Brien K, Pickles T, *et al.* Childhood urinary tract infection in primary care: a prospective observational study of prevalence, diagnosis, treatment, and recovery. *Br J Gen Pract* 2015;**65**:e217-23. doi:10.3399/bjgp15X684361

8 Shaikh N, Ewing AL, Bhatnagar S, *et al.* Risk of Renal Scarring in Children With a First Urinary Tract Infection: A Systematic Review. *Pediatrics* 2010;**126**:1084–91. doi:10.1542/peds.2010-0685

9 Coulthard MG, Lambert HJ, Keir MJ. Occurrence of renal scars in children after their first referral for urinary tract infection. *BMJ* 1997;**315**:918–9.

10 Jacobson SH, Eklöf O, Eriksson CG, *et al.* Development of hypertension and uraemia after pyelonephritis in childhood: 27 year follow up. *BMJ* 1989;**299**:703–6.

11 Round J, Fitzgerald AC, Hulme C, *et al.* Urinary tract infections in children and the risk of ESRF. *Acta Paediatr* 2012;**101**:278–82. doi:10.1111/j.1651-2227.2011.02542.x

12 Smellie JM, Prescod NP, Shaw PJ, *et al.* Childhood reflux and urinary infection: a follow-up of 10-41 years in 226 adults. *Pediatr Nephrol* 1998;**12**:727–36.

13 Craig JC, Williams GJ. Denominators do matter: it's a myth--urinary tract infection does not cause chronic kidney disease. *Pediatrics* 2011;**128**:984–5. doi:10.1542/peds.2011-2631

14 Salo J, Ikäheimo R, Tapiainen T, *et al.* Childhood urinary tract infections as a cause of chronic kidney disease. *Pediatrics* 2011;**128**:840–7. doi:10.1542/peds.2010-3520

15 Coulthard MG, Vernon SJ, Lambert HJ, *et al.* A nurse led education and direct access service for the management of urinary tract infections in children: prospective controlled trial. *BMJ* 2003;**327**:656. doi:10.1136/bmj.327.7416.656

16 Jadresic L, Cartwright K, Cowie N, *et al.* Investigation of urinary tract infection in childhood. *BMJ* 1993;**307**:761–4. doi:10.1136/BMJ.307.6907.761

17 van der Voort J, Edwards A, Roberts R, *et al.* The struggle to diagnose UTI in children under two in primary care. *Fam Pract* 1997;**14**:44–8.

18 Hollingworth W, Busby J, Butler CC, *et al.* The Diagnosis of Urinary Tract Infection in Young Children (DUTY) Study Clinical Rule: Economic Evaluation. *Value Heal* 2017;**20**:556–66. doi:10.1016/j.jval.2017.01.003

19 Jones KH, Ford D V., Jones C, *et al.* A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation. *J Biomed Inform* 2014;**50**:196–204.

doi:10.1016/j.jbi.2014.01.003

20 Ford D V, Jones KH, Verplancke J-P, *et al.* The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;**9**:157. doi:10.1186/1472-6963-9-157

21 Lyons RA, Jones KH, John G, *et al.* The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009;**9**. doi:10.1186/1472-6947-9-3

22 Robling M, Bekkers M-J, Bell K, *et al.* Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial. *Lancet* 2016;**387**:146–55. doi:10.1016/S0140-6736(15)00392-X

23 Public Health England. UK Standards for Microbiology Investigations. Investigation of Urine. B41. 2018. https://www.gov.uk/government/publications/smi-b-41-investigation-of-urine

24 Green SB. How Many Subjects Does It Take To Do A Regression Analysis. *Multivariate Behav Res* 1991;**26**:499–510. doi:10.1207/s15327906mbr2603_7

25 Gilbert R, Lafferty R, Hagger-Johnson G, *et al.* GUILD: GUidance for Information about Linking Data sets†. *J Public Health (Bangkok)* 2017;:1–8. doi:10.1093/pubmed/fdx037

26 von Elm E, Altman DG, Egger M, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;**335**:806–8. doi:10.1136/bmj.39335.541782.AD

27 Benchimol EI, Smeeth L, Guttmann A, *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med* 2015;**12**:e1001885. doi:10.1371/journal.pmed.1001885

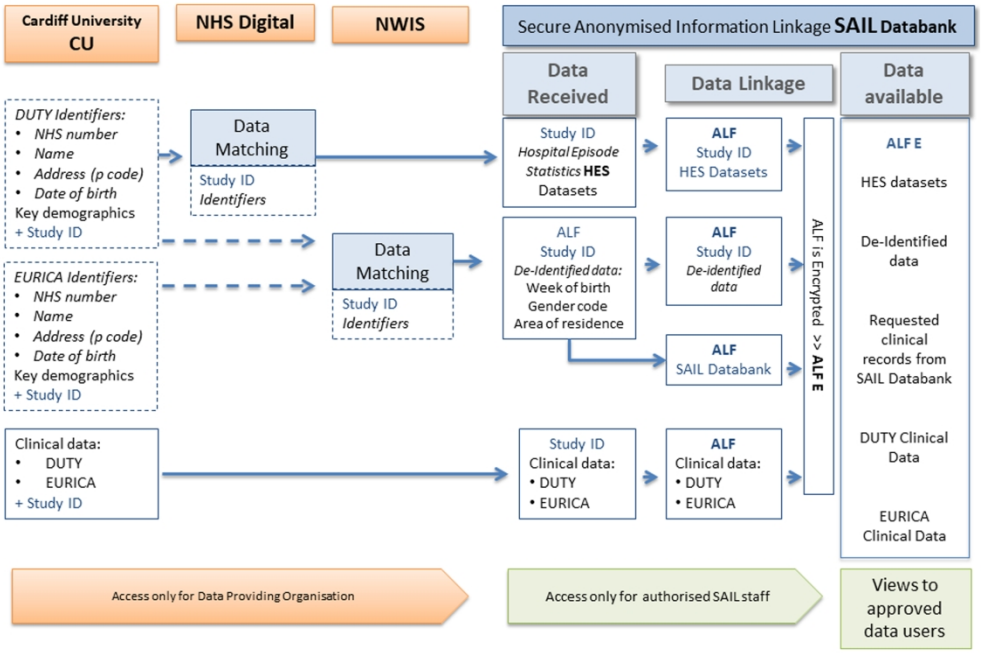
28 IBM. IBM SPSS Statistics for Windows: Version 22.0. 2013.

29 StataCorp. Stata Statistical Software: Release 15. 2017. 2017. doi:10.2307/2234838

(Title) Figure 1. The data flow for dataset 2.

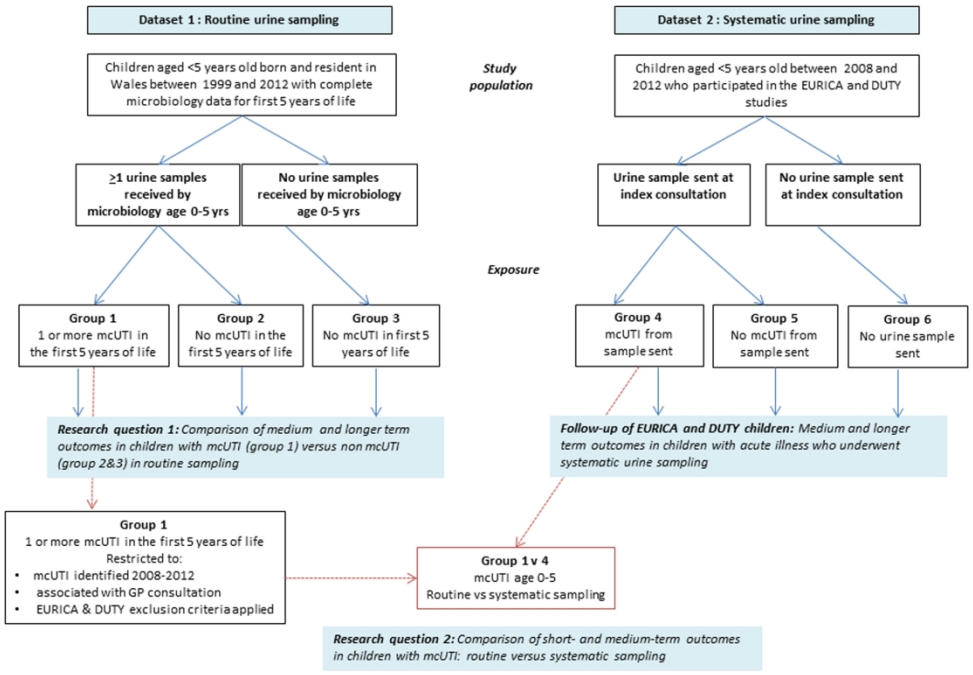
(Legend) ALF- Anonymised Linking Field; NWIS – NHS Wales Informatics Service; HES – Hospital Episode Statistics; SAIL – Secure Anonymised Information Linkage

(Title) Figure 2. Flow chart of study participants



(Title) Figure 1. The data flow for dataset 2.
(Legend) ALF- Anonymised Linking Field; NWIS – NHS Wales Informatics Service; HES – Hospital Episode Statistics; SAIL – Secure Anonymised Information Linkage

297x209mm (300 x 300 DPI)



(Title) Figure 2. Flow chart of study participants

297x209mm (300 x 300 DPI)