BMJ Open

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

Fiona V Lugg-Widger,[1] Lianna Angel,[1] Rebecca Cannings-John,[1] Hywel Jones,[2] Mandy Lau,[1] Christopher Butler,[3] Nick A Francis,[4] Alastair D Hay,[5] Margaret Heginbothom,[6] Kerenza Hood,[1] Shantini Paranjothy,[2] Judith Vandervoort,[7] Kathryn Hughes[4]

ABSTRACT

Introduction Current guidelines advise the prompt diagnosis and treatment of urinary tract infection (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 data sets will be established. The first will consist of routinely collected data (hospital, general practice (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 data set reflects normal practice 'routine sampling'. Outcomes (including renal scarring, hypertension, end-stage renal failure, 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' and '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 an mcUTI in this data set, identified through systematic urine sampling, will be compared with those with an mcUTI identified through routine urine sampling (data set 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 multisourced data set will allow a comparison of outcomes over 5 years for children with and without microbiologically confirmed urinary tract infection 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 prioritise interventions to improve early diagnosis, sampling and treatment, potentially improving health outcomes and reducing National Health Service 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 5 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 risks 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 Care 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 (eg, 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 5 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 microbiologically confirmed UTI (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 mcUTIs aged less than 5 years old, compared with those with no mcUTI, using National Health Service (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 5 years following participation in two UK prospective cohort studies of acutely ill children with systematic urine sampling in primary care, the Diagnosis of Urinary Tract infection in Young children (DUTY) and EURICA (a Welsh cohort study of urinary tract infection in children) 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 Long-term outcomes of UTI in Childhood (LUCI) study will use data linkage of routinely collected data sets 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 with children who have not experienced an 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 with routine sampling (standard, clinician-led sampling)?

Study design
This is a data linkage study comprising two overarching data sets of children. Data set 1 will comprise routinely collected health data from children born and resident in Wales. Children in this data set will have had urine sampled according to routine practice. Routine data will be available on all children for 7 years, and longer for some (ie, children will be followed up until the date of data abstraction). Data set 1 will be used to answer research question 1.

Data set 2 will be children who participated in the EURICA or DUTY studies. Children in this data set 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 Secure Anonymised Information Linkage, SAIL). Follow-up will be for a minimum of 5 years, and longer where data are available. Data set 2 will be used to describe longer term follow-up for DUTY and EURICA study children. Those with mcUTI in data set 2 will be compared with those with mcUTI identified in data set 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.

Data providers and data sets
The EURICA and DUTY studies
This work builds on two large cohort studies of acutely ill children, aged less than 5 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

Table 1 Sources of data

Data provider	Data source	Dates available	Indicative/key data items	Data set 1 (routine sampling)	Data set 2 (systematic sampling)
SAIL (Wales)	GP	January 1994 to October 2016	Secondary outcomes including antibiotic prescriptions, GP consultations, chronic kidney disease, hypertension	✓	✓
	Patient Episode Database for Wales (PEDW)	January 1994 to April 2017	Primary and secondary outcomes and covariates including renal scarring, hospital admission, end-stage renal failure, VUR, renal/bladder surgery	✓	✓
	Welsh Demographic Service (WDS)	January 1994 to April 2017	Demographics	✓	✓
	Welsh Electronic Cohort of Children (WECC)	January 1994 to September 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	January 1994 to April 2017	Primary outcome—renal scarring	✓	✓
Individual Health Boards	Radiology data	January 1994 to January 2017	Validation of the primary outcome (renal scarring)	✓	
NHS Digital (England)	Admitted patient care	April 2008 to March 2017	Primary and secondary outcomes and covariates including renal scarring, hospital admission, end-stage renal failure, VUR, renal/bladder surgery		✓
	Outpatient	April 2008 to March 2017	Primary outcome—renal scarring		✓
Bristol University (England and Wales)	DUTY	April 2010 to April 2012	Baseline characteristics, UTI status, presenting symptoms and signs, initial clinical management		✓
Cardiff University (Wales)	EURICA	July 2008 to July 2010	Baseline characteristics, UTI status, presenting symptoms and signs, initial clinical management		✓

DUTY, Diagnosis of Urinary Tract infection in Young children; EURICA, Welsh cohort study of urinary tract infection in children; GP, general practice; SAIL, Secure Anonymised Information Linkage; UTI, urinary tract infection; VUR, vesicoureteric reflux.

study and analysed in NHS microbiology laboratories. In the EURICA study, 1003 children were recruited from general practitioner's (GP) practices in Wales between March 2008 and July 2010. 5.9 per cent of 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 the National Institute for Social Care

and Health Research (now Health and Care Research Wales) and sponsored by Cardiff University, and DUTY was funded by the National Institute for Health Research-Health Technology Assessment and sponsored by Bristol University.

SAIL Databank

The study will use the 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 data sets 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. This study will access Hospital Episode Statistics (HES) data for participants of the DUTY and EURICA studies. 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 include diagnosis, procedures and length of episode according to the codes of the 10th Revision of the International Statistical Classification of Diseases and Related Health Problems.

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 data set 1. A one off data extract of patient-level attendance data for patients born between 1 January 1994 and 31 December 2012 who attended radiology between 1994 and 2016 will be transferred to SAIL. Data extracted include examination performed, attendance data and the radiology report.

Opportunity to opt out (data set 2)

Data set 1 uses routinely collected data that are fully anonymised so we do not require individual consent in order to access these data. Data set 2 involves participants from the DUTY and EURICA studies and therefore requires section 251 (s251) support of the 2006 NHS Act approval from the Health Research Authority's Confidentiality Advisory Group (CAG) to pass identifiable participant data legally held by Cardiff University 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 data sets sent to the information centres and therefore will not appear in the data sets 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 NHS Wales Informatics Service (NWIS), their trusted third party) 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 and postcode. Data matching by NHS Digital will be separate from 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 data set

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 data sets will be transferred to SAIL following a process of deidentification. The data flow is shown in figure 1. The study ID will be replaced by an anonymised linking field (ALF) to enable all data sets to be linked at an individual level. The resulting data set will contain clinical variables (from DUTY, EURICA, SAIL and NHS Digital), deidentified demographic variables (eg, week of birth and lower super output area) and the ALF. The key between study ID and ALF will be retained and encrypted as a further safeguard. The ALF is further encrypted so that data sets and records cannot be linked across multiple projects a researcher has access to.
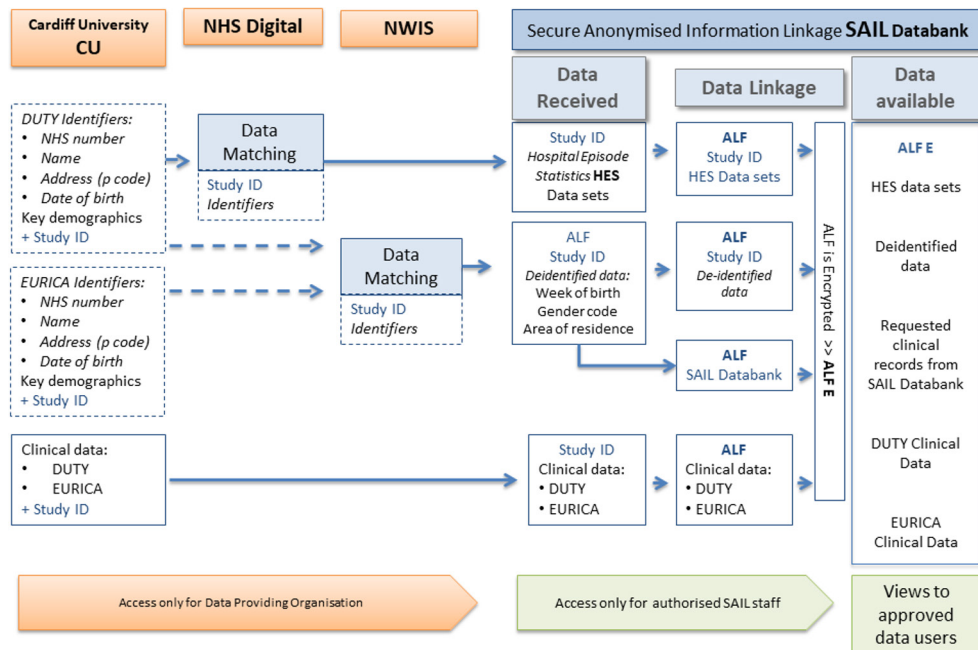
Figure 1 The data flow for data set 2. ALF, anonymised linking field; DUTY, Diagnosis of Urinary Tract infection in Young children; EURICA, Welsh cohort study of Urinary Tract Infection in Children; NHS, National Health Service; NWIS, NHS Wales Informatics Service.

Study participants

A flow chart of participants in the study is shown in figure 2. Data set 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 1 January 2005 and 31 December 2009 to ensure that all

of the first 5 years of life are covered by the dates when Datastore is available.

Data set 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
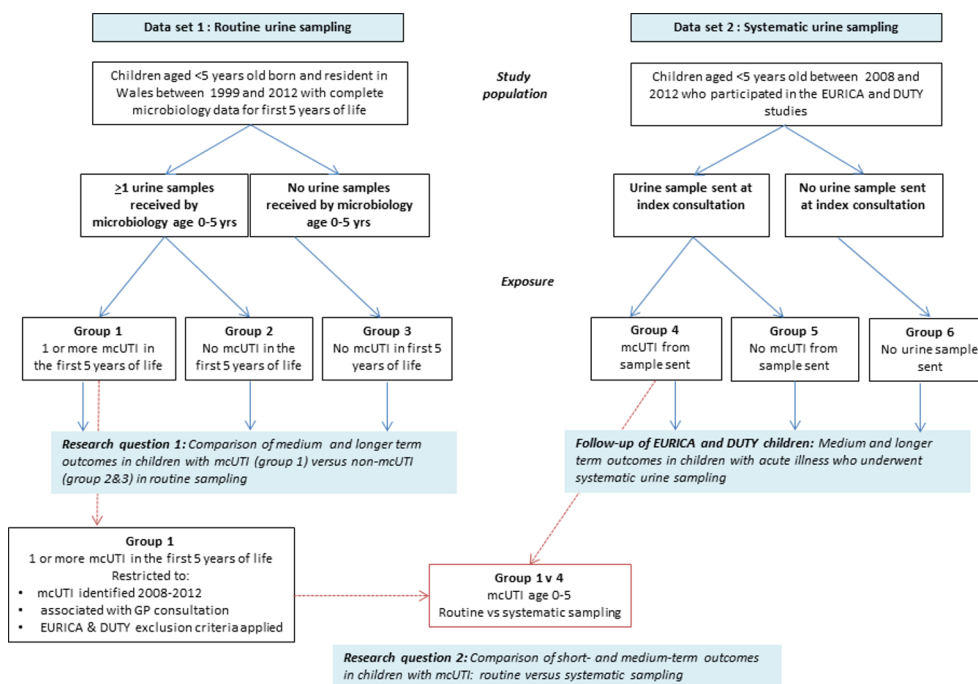
Figure 2 Flow chart of study participants. DUTY, Diagnosis of Urinary Tract infection in Young children; GP, general practice; mcUTI, microbiologically confirmed urinary tract infection.

when comparing children with an mcUTI from data sets 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. Data sets 1 and 2 will be limited to children with an mcUTI associated with a GP consultation between March 2008 and April 2012. For both data sets, 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 (ie, 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 data set 1 to make them directly comparable to data set 2; therefore, children will be excluded from the routine sampling cohort if any criterion met:

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

Children with a 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 data set 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 in hospital samples.[6] NHS laboratories take into consideration the nature of the urine sample in their reporting.[23] For data set 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 data set 2, the presence of significant bacteriuria (pure or predominant growth of 100 000 cfu/mL of urine) will be used to define UTI.

For data set 1 we define the exposure period as <5 years and will be grouped as follows (figure 2):

Group 1: children with at least one mcUTI before their fifth birthday or before outcome of interest.
Group 2: children with at least one urine sample but no mcUTI before their fifth birthday or before outcome of interest.
Group 3: children with no urine samples before their fifth 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 fifth birthday will be taken.

For the main analyses, groups 2 and 3 will be considered together as having no mcUTI. Data set 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 an 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 Welsh Demographic Service, Welsh Electronic Cohort of Children and for a subset with GP records. The study outcomes are summarised in table 3.

Follow-up

Follow-up of all children in data sets 1 and 2 will continue until the first occurrence of any of: outcome, migration, death or end of follow-up; and for the subanalysis 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 5 years of life were covered by the dates that Datastore was available (excluding children whose first 5 years of life fall outside of Datastore availability) and limiting date of births to 1 January 2005 to 31 December 2009, will mean that the shortest follow-up period will be 7 years. Follow-up will be longer where data are 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 OR of 1.5 (a conservative effect size of

Table 2 Child and maternal characteristics

Risk factor	Data set 1: routine sampling Source	Data set 2: systematic sampling Source
Age of child at first urine sample (group 1/2 only) or index consultation (data set 2)	Datastore	EURICA and DUTY study data
Gender	WECC	EURICA and DUTY study data
Ethnicity	WECC	WECC and DUTY study data
Deprivation quintile at birth (taken from postcode at birth)	WDS: Welsh index of multiple deprivation and Townsend score	Townsend score; EURICA and DUTY study data
Maternal age at birth (years) (category)	WECC	EURICA and DUTY Welsh participants only: WECC
Birth weight (g) (category)	WECC	EURICA and DUTY Welsh participants only: WECC
Gestational age at birth (weeks) (category)	WECC	EURICA and DUTY Welsh participants only: WECC
Ever breast fed	WECC	EURICA and DUTY study data
Uropathogen (Enterobacteriaceae or other) (mcUTI only)	Datastore	EURICA and DUTY study data
Antimicrobial resistance (mcUTI only)	Datastore	EURICA and DUTY study data
Congenital malformations (con mals)		
Known to be associated with UTI/renal scarring	PEDW	
1. Spina bifida/neurobladder		NA: DUTY exclusion
2. Renal/urinary system con mals including vesicoureteric reflux (VUR)		EURICA and DUTY study data
Possibly associated with UTI/renal scarring		
1. Down's syndrome	PEDW; WECC	EURICA and DUTY study data
2. Cerebral palsy/other paralytic syndromes	WECC	EURICA and DUTY study data
Congenital malformations—major/minor	WECC	EURICA and DUTY Welsh participants only: WECC
Comorbidities		
1. Diabetes diagnosed under the age of 5 years	PEDW	EURICA and DUTY study data*
2. Renal or urogenital surgery	PEDW	EURICA study data* DUTY exclusion
3. Cancer	PEDW	EURICA and DUTY study data*
4. Immunosuppressive disease	PEDW	EURICA and DUTY study data*
5. Circumcision (aged <5 years)	PEDW; GP	EURICA and DUTY study data*
Factors for follow-up of study participants and research question 2		
Symptoms and signs at index consultation	–	EURICA and DUTY study data
Management at index consultation	GP	EURICA and DUTY study data
Antenatal ultrasound urinary system abnormalities	–	EURICA and DUTY study data
Family history of UTI/urinary system problems	–	EURICA and DUTY study data
Recent antibiotics (7 days prior to index consultation)	GP	EURICA and DUTY* study data

*At time of index consultation.
DUTY, Diagnosis of Urinary Tract infection in Young children; EURICA, Welsh cohort study of urinary tract infection in children; GP, general practice; mcUTI, microbiologically confirmed UTI; NA, not applicable; PEDW, Patient Episode Database for Wales; UTI, urinary tract infection; WDS, Welsh Demographic Service; WECC, Welsh Electronic Cohort of Children.

5% difference between groups), two-sided alpha 0.05, power=90% and given a ratio of 32:1 (children diagnosed with mcUTI:children with no UTI diagnosis) in the data set, we require 15 519 children (452 with mcUTI:15 067 without mcUTI) for analysis. Initial examination of the SAIL data set identified just under 13 000 children less than 5 years old with UTI between 1999 and 2012. As our sample size calculation requires 452 with

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 (5–7 years follow-up)			✓

GP, general practice; PEDW, Patient Episode Database for Wales; UTI, urinary tract infection.

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 *et al*.[8] Assuming a proportion of 8% with renal scarring in the mcUTI group, and an OR of 1.5 (2.5% difference), two-sided alpha=0.05, power=90% and a ratio of 32:1, a total of 35 833 children (1044 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 mcUTI 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 two-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 (OR=1.5), a two-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
Data set 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 (IQR)) to summarise the population for the main analyses by groups 1, 2 and 3 and any marked imbalance between the groups will be identified. The main comparative analyses will be carried out at a child level since outcomes relate to an individual's exposure to one or more UTIs and will test the null hypothesis that there are no differences in outcomes in the first 5 years and between the ages 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% 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 HRs with 95% CIs for each exposure group.

We will construct a directed acyclic graph 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 a 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 subgroup analyses for the primary outcome: gender of child and presence of any renal/urological congenital

anomalies. These preplanned 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 (eg, 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 and resistance) in children presenting with a childhood mcUTI. In a subsample 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 an 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.

Data set 2: systematic sampling of UTI

Detailed study data for DUTY and EURICA participants are 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 inpatient 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 after 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 Patient Episode Database for Wales 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 subset. 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 versus selective sampling. Predictors of outcome will be examined as before using a multilevel multinomial regression model (no event, 30 days, event 30 days to 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 first 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 these 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-randomised evaluation of a public health intervention. SPSS (Version 25) and Stata (StataCorp, 2017) will be used for all analyses.[28]

Patient and public involvement

We have a parent representative (Sarah Jones) who has contributed to all stages of this study. She helped 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 data set 1 differs from data set 2. Data set 1 is an anonymised data set made available from SAIL Databank with only approval required from the IGRP whereas data set 2 involves the transfer of identifiable data to data providers which requires ethical approval, s251 support, IGRP and IGARD approval. In order to obtain an unbiased sample from the DUTY and EURICA cohorts, 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 data sets will give us a large cohort including demographic, hospital inpatient and outpatient, 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 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 (1) this protocol paper, (2) main results from research question 1, and (3) main results from research question 2. The findings from this study will be of interest to clinicians and policymakers and may influence the management of acutely ill children and childhood UTI.

Author affiliations
[1]Centre for Trials Research, Cardiff University, Cardiff, UK
[2]Division of Population Medicine, National Centre for Population Health and Wellbeing Research, Cardiff University, Cardiff, UK
[3]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
[4]Division of Population Medicine, School of Medicine, Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales), Cardiff University, Cardiff, UK
[5]Centre for Academic Primary Care, University of Bristol, Bristol, UK
[6]Microbiology, Public Health Wales, Cardiff, UK
[7]Paediatric Nephrology, Noah's Ark Children's Hospital for Wales, Cardiff, UK

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.
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;113:5S–13.
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* 2016;20:1–294.
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.
6. NICE. Urinary tract infection in under 16s: diagnosis and management: Guidance and guidelines.
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.
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.
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.
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.
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.
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.
16. Jadresic L, Cartwright K, Cowie N, *et al*. Investigation of urinary tract infection in childhood. *BMJ* 1993;307:761–4.

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 in Health* 2017;20:556–66.
19. Jones KH, Ford DV, 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.
20. Ford DV, Jones KH, Verplancke JP, *et al*. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;9:157.
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.
22. Robling M, Bekkers MJ, 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.
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.
25. Gilbert R, Lafferty R, Hagger-Johnson G, *et al*. GUILD: GUidance for Information about Linking Data sets. *J Public Health* 2018;40:191–8.
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.
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.
28. IBM. IBM SPSS Statistics for Windows: Version 22.0.. 2013.