

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email editorial.bmjopen@bmj.com

BMJ Open

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

Journal:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 1 Long-term outcomes of urinary tract infection (UTI) in Childhood (LUCI): protocol for an
4 2 electronic record-linked cohort study
5
6 3

7 4 **Corresponding author:**

8 5 Fiona Lugg-Widger LuggFV@cardiff.ac.uk

9 6 Address: 4th Floor Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4YS.
10

11 7
12 8 Lianna Angel¹

AngellL@cardiff.ac.uk

13 9 Rebecca Cannings-John¹

CanningsRL@cardiff.ac.uk

14 10 Hywel Jones²

JonesH75@cardiff.ac.uk

15 11 Mandy Lau¹

LauTM@cardiff.ac.uk

16 12 Christopher C Butler³

christopher.butler@phc.ox.ac.uk

17 13 Nick Francis⁴

FrancisNA@cardiff.ac.uk

18 14 Alastair D Hay⁵

alastair.hay@bristol.ac.uk

19 15 Margaret Heginbothom⁶

margaret.heginbothom@wales.nhs.uk

20 16 Kerenza Hood¹

HoodK1@cardiff.ac.uk

21 17 Shantini Paranjothy²

ParanjothyS@cardiff.ac.uk

22 18 Judith Van der Voort⁷

judith.vandervoort@wales.nhs.uk

23 19 Kathryn Hughes⁴

HughesKA6@cardiff.ac.uk
24
25

26 21 **Author Affiliations**

- 27 22 1. Centre for Trials Research, College of Biomedical & Life Sciences, Cardiff University,
28 23 Cardiff, UK.
29 24 2. National Centre for Population Health and Wellbeing Research, Division of
30 25 Population Medicine, Cardiff University
31 26 3. Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe
32 27 Primary Care Building, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2
33 28 6GG
34 29 4. Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales),
35 30 Division of Population Medicine, School of Medicine, Cardiff University
36 31 5. Centre of Academic Primary Care, NIHR School for Primary Care Research, Bristol
37 32 Medical School: Population Health Sciences, Canynge Hall, 39 Whatley Road, Bristol
38 33 BS8 2PS
39 34 6. Public Health Wales, 2 Capital Quarter, Tyndall Street, Cardiff CF10 4BZ
40 35 7. Noah's Arc Childrens Hospital for Wales, Cardiff, CF14 4XW
41
42
43
44
45
46
47
48

49 39 **Word Count:** 6382

50
51 40 **Keywords:** Children; Urinary Tract Infection; Medical record linkage; Renal scarring; Urine
52 41 sampling; Primary care.
53
54 42

55 43 **ABSTRACT**
56
57
58
59
60

1
2
3 44 **Introduction:** Current guidelines advise the prompt diagnosis and treatment of UTI in
4
5 45 children to improve both short and longer term outcomes. However, the risk of long-term
6
7 46 complications following childhood UTI is unclear.

8
9 47 UTI is relatively common but difficult to diagnose in children as symptoms are non-specific.
10
11 48 Diagnosis requires a urine sample, but sampling is difficult and infrequent, and it is not clear
12
13 49 if sampling should be given greater priority in primary care. The LUCI study will assess the
14
15 50 short, medium and longer-term outcomes of childhood UTI associated with routine and
16
17 51 systematic sampling practices.

18
19
20 52 **Methods and analysis:** Two datasets will be established: The first will consist of routinely
21
22 53 collected data (Hospital, GP, Microbiology) from children born and resident in Wales, linked
23
24 54 via the Secure Anonymised Information Linkage (SAIL) databank (an 'e-cohort'). Urine
25
26 55 sampling in this dataset reflects normal practice 'routine sampling'. Outcomes (including
27
28 56 renal scarring, hypertension, end-stage renal failure(ESRF), hospital admissions, GP
29
30 57 consultations, antibiotic prescriptions) for children with at least one UTI confirmed with
31
32 58 microbiological culture (mcUTI) or no mcUTI before the age of 5 will be compared.

33
34 59 The second will combine data from two prospective observational studies('DUTY' &
35
36 60 'EURICA') employing systematic urine sampling for children presenting to primary care with
37
38 61 acute, undifferentiated illness, linked to routine data via SAIL (Wales) and NHS Digital
39
40 62 (England). Outcomes (as above, plus features of mcUTI) for children with a mcUTI in this
41
42 63 dataset, identified through systematic urine sampling, will be compared to those with a
43
44 64 mcUTI identified through routine urine sampling (dataset 1).

45
46
47 65 **Ethics and dissemination:** The study protocol has been approved by NHS Wales
48
49 66 Research Ethics Committee and the Health Research Authority's Confidentiality Advisory
50
51 67 Group. Methods of innovative study design and findings will be disseminated through peer-
52
53 68 review journals and conferences. Results will be of interest to clinical and policy
54
55 69 stakeholders in the UK.

70 **Strengths and limitations of this study:**

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

85 **INTRODUCTION**

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

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

1
2
3 96 were conducted in secondary care and most required fever for inclusion in the study [8].

4 97 These are likely to represent more serious UTIs than those presenting in primary care, and

5 98 this rate of renal scarring may not apply to all children presenting with UTI. The risk of long-

6 99 term complications (such as hypertension and renal failure) as a result of renal scarring

7
8
9
10 following childhood UTI is also not clear, with some researchers questioning this association.

11
12 101 [6,13,14] A review, commissioned by the National Institute for Health and Clinical Excellence

13
14 102 (NICE), concluded that 'there are no appropriate studies that accurately estimate the risks of

15
16 103 long term complications as a result of childhood UTI'. [6]

17
18
19 104 A urine sample is necessary to confirm the presence of UTI in childhood as the presenting

20
21 105 symptoms are non-specific and similar to those found in many common childhood illnesses.

22
23 106 [6] Furthermore, microbiological confirmation is important as some children with UTI require

24
25 107 invasive investigations. [6] Children who become seriously ill and who are assessed in

26
27 108 emergency departments or admitted to hospital will usually have their urine sampled. [1]

28
29 109 However, in primary care, where most acute illness is seen, urine is infrequently sampled. It

30
31 110 is estimated that urine is sampled in fewer than 2% of acute illness consultations with

32
33 111 children under five years old in the UK. [4] Studies have suggested that many cases of

34
35 112 childhood UTI are missed in primary care. [15–17] Routine practice (urine sampling based

36
37 113 on clinician suspicion) is likely to miss more than two thirds of UTIs in primary care. [5,7] The

38
39 114 clinical implications of missed childhood UTIs are not known. Increasing urine sampling in

40
41 115 primary care is likely to increase the diagnosis of childhood mcUTI, and is advocated by

42
43 116 current guidelines, but it is not clear whether this is an appropriate strategy, whether

44
45 117 outcomes for children would improve or to what extent it should be increased. [4–6,15,18]

46
47 118 Clarification is needed on two issues: First, to determine what the longer-term outcomes are

48
49 119 following childhood UTI (including those identified in primary care), and second, to determine

50
51 120 whether outcomes vary according to sampling practice.

52
53
54 121 In this study we will describe clinical outcomes for all children with one or more mcUTI aged

55
56 122 less than five years old, compared to those with no mcUTI, using NHS laboratory data from

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

125 We will also describe longer-term follow up of clinical outcomes (including renal scarring) for
126 at least five years following participation in two UK prospective cohort studies of acutely ill
127 children with systematic urine sampling in primary care, the DUTY and EURICA studies.
128 [5,7] We will compare the outcomes of those with mcUTI identified through these studies
129 (systematic urine sampling) with those identified through routine practice.

130 **METHODS AND DESIGN**

131 **Research objectives**

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

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

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

140 **Study Design**

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

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

1
2
3 148 were asked to provide a urine sample). DUTY and EURICA children will be followed-up by
4
5 149 linking records to routinely collected health data from England (using NHS Digital) and
6
7 150 Wales (using SAIL). Follow-up will be for a minimum of five years, and longer where data is
8
9 151 available. Dataset 2 will be used to describe longer term follow-up for DUTY and EURICA
10
11 152 study children. Those with mcUTI in Dataset 2 will be compared to those with mcUTI
12
13 153 identified in Dataset 1 to answer Research Question 2.

14
15 154 The study formally started in October 2016 and will report to funder in October 2018. A
16
17 155 summary of the data sources is provided in Table 1.

18
19
20 156

21 157

22
23 158
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

For peer review only

1
2
3 161 **Data providers and datasets**
4

5 162 The EURICA and DUTY Studies
6
7

8 163 This work builds on two large cohort studies of acutely ill children, aged less than five years
9
10 164 old, presenting in primary care, in which mcUTI status was determined using systematic
11
12 165 urine sampling [5,7]. In both studies, clinical and demographic data were collected and urine
13
14 166 samples requested from all children included in the study and analysed in NHS microbiology
15
16 167 laboratories. In the EURICA study, 1003 children were recruited from GP practices in Wales
17
18 168 between March 2008 and July 2010. 5.9% children had mcUTI. Participants had a GP notes
19
20 169 review at 6 months. In the DUTY study, 7163 children were recruited from GP practices in
21
22 170 England and Wales between April 2010 and April 2012 and 5.6% children had mcUTI
23
24 171 confirmed in NHS labs. An algorithm to predict UTI based on presenting symptoms and
25
26 172 signs was developed. Neither study had sufficient follow-up to determine whether renal
27
28 173 investigations to look for renal scarring had been undertaken or found. EURICA was funded
29
30 174 by NISCHR (now Health and Care Research Wales) and sponsored by Cardiff University
31
32 175 and DUTY was NIHR HTA funded and sponsored by Bristol University.
33

34 176
35

36 177 SAIL Databank
37

38
39 178 The study will use the Secure Anonymised Information Linkage (SAIL) Databank to access
40
41 179 routinely collected data outlined in Table 1. SAIL is a repository for a broad range of routinely
42
43 180 collected health and population data in Wales. SAIL will also act as a data safe haven for the
44
45 181 clinical DUTY and EURICA datasets and data made available from NHS Digital and
46
47 182 Individual Health Boards. All data will be accessed via the SAIL Gateway following
48
49 183 Information Governance Review Panel (IGRP) approval. SAIL Databank does not handle
50
51 184 any identifiable data therefore all data will be anonymised including data transferred from
52
53 185 other information centres [19–21].
54

55 186
56
57
58
59
60

1
2
3 187 NHS Digital

4
5 188 NHS Digital is the trading name of the Health and Social Care Information Centre (HSCIC).
6
7 189 This study will access Hospital Episode Statistics (HES) data for participants of the DUTY
8
9 190 and EURICA study. All available Inpatient and Outpatient records belonging to each study
10
11 191 participant will be requested and approved by the Independent Group Advising on the
12
13 192 Release of Data (IGARD) Panel. The data requested includes diagnosis, procedures and
14
15 193 length of episode according to the 10th revision of the International Statistical Classification
16
17 194 of Diseases and Related Health Problems [ICD-10] codes.
18

19
20 195

21
22 196 Public Health Wales

23
24
25 197 Public Health Wales will provide a data extract of urine microbiology culture results from all
26
27 198 microbiology laboratories in Wales (Datastore) for use with this project. This will be
28
29 199 transferred to SAIL.
30

31 200

32
33 201 Individual Health Boards

34
35 202 Health boards in Wales will be approached to access anonymised radiology data for patients
36
37 203 in dataset 1. A one off data extract of patient-level attendance data for patients born
38
39 204 between 01/01/1994 and 31/12/2012 that attended radiology between 1994 and 2016 will be
40
41 205 transferred to SAIL. Data extracted includes examination performed, attendance data and
42
43 206 the radiology report.
44

45
46 207

47
48
49 208 **Opportunity to opt-out (dataset two)**

50
51 209 Dataset one uses routinely collected data that is fully anonymised so we do not require
52
53 210 individual consent in order to access these data. Dataset two involves participants from the
54
55 211 DUTY and EURICA study and therefore requires section 251 (s251) support of the 2006
56
57

1
2
3 212 NHS Act approval from the Health Research Authority's Confidentiality Advisory Group (HRA
4 213 CAG) to pass identifiable participant data legally held by Cardiff and Bristol University (as
5 214 sponsors of the studies) to the data providers. This is using an opt-out/dissent model instead
6
7 215 of obtaining further consent. In order to provide the opportunity for participants to dissent a
8
9 216 letter has been developed to be sent to all parents of participants. Participants have the
10
11 217 opportunity to contact the study team through an online web-form, by email, text or
12
13 218 telephone (details provided on website) and register their dissent. These participants who
14
15 219 register their dissent will not be included in any of the datasets sent to the information
16
17 220 centres and therefore will not appear in the datasets for analysis. A participant
18
19 221 representative was consulted on the layout, wording and level of information contained in the
20
21 222 participant opt-out letter and on the study website. A key consideration was to communicate
22
23 223 the data transfer process. The final letter was approved by both an NHS Research Ethics
24
25 224 Committee and CAG committee as part of overall governance approval for the study.
26
27
28

29
30
31

32 **Data Matching**

33
34 227 For included data sets, data will be sent to NHS Digital and SAIL (via NWIS their trusted
35
36 228 third party (TTP)) for matching using a combination of NHS number, name, address and
37
38 229 date of birth. Matching with NHS Digital data will be by exact matching on NHS Number,
39
40 230 Date of Birth and Postcode. This has been conducted in other studies and achieved a high
41
42 231 match rate (99.6%) [22]. NWIS match using NHS Number, Date of Birth, Sex, Postcode.
43
44 232 Data matching by NHS Digital will be separate to matching conducted by NWIS on behalf of
45
46 233 SAIL. We would expect only a small number of participants matched to both English and
47
48 234 Welsh NHS records however there is the possibility of this for those using services across
49
50 235 the border to their current address.
51

52
53
54

55 **The anonymised dataset**

1
2
3 238 For DUTY and EURICA participants NHS Digital will retain the unique study ID assigned by
4
5 239 the DUTY/EURICA study, with the HES records after the matching process before it is sent
6
7 240 to SAIL Databank. This ID will be retained for data transfer to ensure the data can be linked.
8
9 241 The same applies to data sent to NWIS. Requested data will then be linked for each
10
11 242 participant via SAIL using the unique study ID. The DUTY and EURICA clinical datasets will
12
13 243 be transferred to SAIL following a process of de-identification. The data flow is shown in
14
15 244 Figure 1. The study ID will be replaced by an anonymised linking field (ALF) to enable all
16
17 245 datasets to be linked at an individual level. The resulting dataset will contain clinical
18
19 246 variables (from DUTY, EURICA, SAIL and NHS Digital), de-identified demographic variables
20
21 247 (e.g. week of birth and lower super output area - LSOA) and the ALF. The key between
22
23 248 study ID and ALF will be retained and encrypted as a further safeguard. The ALF is further
24
25 249 encrypted (ALF-E) so that datasets and records cannot be linked across multiple projects a
26
27 250 researcher has access to.
28
29 251

31 252 **Study Participants**

32
33
34 253 A flow chart of participants in the study is shown in Figure 2. Dataset 1 (routine sampling)
35
36 254 will be children born and resident in Wales for the first 5 years of life; who were less than 5
37
38 255 years old between 1999 and 2012. The main analysis will be on children born between
39
40 256 01/01/2005 and 31/12/2009 to ensure that all of the first five years of life are covered by the
41
42 257 dates when Datastore is available.
43

44 258 Dataset 2 (systematic sampling from DUTY and EURICA) will be children who participated in
45
46 259 the EURICA or DUTY studies, who were not withdrawn from the study and when provided
47
48 260 with the opportunity to opt-out, remained in the study. For research question 2 when
49
50 261 comparing children with a mcUTI from Dataset 1 and 2, children in Group 1 will be selected
51
52 262 to match the DUTY and EURICA study eligibility criteria as closely as possible within the
53
54 263 constraints of using routine data. Datasets 1 and 2 will be limited to children with a mcUTI
55
56 264 associated with a GP consultation between March 2008 and April 2012. For both datasets,

1
2
3 265 the child's mcUTI, associated with a GP consultation (defined as within 14 days prior to the
4
5 266 sample) will be identified and defined as the index consultation. To limit the potential
6
7 267 transfer of GP systematic sampling behaviour, children from Group 1 with index
8
9 268 consultations between 2008 and 2012 at practices which participated in the EURICA study
10
11 269 will be excluded as will children with index consultations between 2010 and 2012 at
12
13 270 practices which participated in the DUTY study. Children will only be included once in each
14
15 271 study period (i.e. a child with a sample sent within the EURICA study period could also
16
17 272 appear in the DUTY study period). In addition, we will apply the DUTY study exclusion
18
19 273 criteria (where possible) to Dataset 1 to make them directly comparable with Dataset 2;
20
21 274 therefore children will be excluded from the routine sampling cohort if any criterion met:
22

275

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

280

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

284

285
286

287 **Exposure**

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

1
2
3 293 participants were from England (Datastore is Wales only). For Dataset 2, the presence of
4
5 294 significant bacteriuria (pure or predominant growth of 100,000 cfu/ml of urine) will be used to
6
7 295 define UTI.

8
9 296 Dataset 1 will be divided into three groups based on the first five years of life (Figure 1).

10
11
12 297 Group 1: children with at least one mcUTI

13
14 298 Group 2: children with at least one urine sample but no mcUTI

15
16
17 299 Group 3: children with no urine samples

18
19
20 300 Exposure is a discrete time-varying covariate and will be taken at the point of outcome;
21
22 301 otherwise the exposure status of the child at their 5th birthday will be taken. For the main
23
24 302 analyses, Groups 2 and 3 will be considered together as having no microbiologically
25
26 303 confirmed UTI. Dataset 2 will similarly be divided into three groups based on their index
27
28 304 consultation when recruited into the DUTY and EURICA studies:

29
30 305 Group 4: children with a mcUTI

31
32 306 Group 5: children who had a urine sample but no mcUTI

33
34 307 Group 6: children who had no urine sample

35
36
37
38 308

39
40 309

41 310 **Study variables**

42
43
44 311 Table 2 shows a breakdown of the baseline data and possible covariates available for
45
46 312 children and maternal characteristics from the data collection forms for EURICA and DUTY
47
48 313 and WDS, WECC, and for a subset with GP records. The study outcomes are summarised
49
50 314 in Table 3.

51
52 315

53
54 316

55
56 317

318 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)		
<i>known to be associated</i> 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
<i>possibly associated</i> 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

319 * at time of index consultation

320 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)			✓

321

322 **Follow-up**

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

331

332 **Analysis**

333 Sample size

334 *Comparison of renal scarring in children routinely sampled for UTI with and without mcUTI:*

335 The sample size is based on the outcome of renal scarring of children with and without

1
2
3 336 mcUTI and taken as 15% [8]. Using an odds ratio of 1.5 (a conservative effect size of 5%
4
5 337 difference between groups), 2-sided alpha 0.05, power=90%, and given a ratio of 32:1
6
7 338 (children diagnosed with mc UTI: children with no UTI diagnosis) in the dataset, we require
8
9 339 15,519 children (452 with mcUTI:15,067 without mcUTI) for analysis. Initial examination of
10
11 340 the SAIL dataset identified just under 13,000 children less than five years old with UTI
12
13 341 between 1999 and 2012. As our sample size calculation requires 452 with UTI, we can be
14
15 342 confident of adequate power for this study. However, the true proportion with renal scarring
16
17 343 is likely to be less than that reported by Shaikh [8]. Assuming a proportion of 8% with renal
18
19 344 scarring in the mcUTI group, and an odds ratio of 1.5 (2.5% difference), 2-sided alpha=0.05,
20
21 345 power=90%, and a ratio of 32:1, 35,833 children (1,044 with UTI: 34,789 without UTI) are
22
23 346 required for analysis, which is still achievable.

24
25 347 *Comparison of systematically versus routinely sampled UTI:* This sample size is constrained
26
27 348 by the number of children with a systematically sampled microbiologically confirmed UTI by
28
29 349 NHS lab (N=374). Using the lower estimate of renal scarring of 8% of routinely sampled
30
31 350 UTIs, then a 5% difference (13% in the systematically sampled group) would give 89%
32
33 351 power, with a 2-sided alpha=0.05 and a ratio of 32:1 (routinely sampled UTIs 11,968:
34
35 352 systematically sampled UTIs 374).

36
37
38 353 Both analyses use multivariable regression. Using Green's [23] formulae, assuming medium
39
40 354 effect sizes (odds ratio=1.5), a 2-sided alpha=0.05, and power=80%, to test 20 predictors in
41
42 355 the multivariable regression model we require at least 159 children in total. This suggests
43
44 356 that we will be adequately powered for both analyses (given these assumptions) to examine
45
46 357 predictors of short and medium-term outcomes.

47
48 358

49
50
51 359 Statistical analysis

52
53 360 *Dataset 1: Routine sampling of UTI*
54
55
56
57
58
59
60

1
2
3 361 Research question 1: Comparison of medium (up to 5 years) and longer term (≥ 5 years)
4 362 outcomes in children with mcUTI versus those without mcUTI in routine sampling.

5
6
7 363 Baseline variables will be described using appropriate descriptive summaries (N (%), mean
8 364 (SD), median (interquartile range)) to summarise the population for the main analyses by
9
10
11 365 group 1, 2 and 3 and any marked imbalance between the groups will be identified. There will
12
13 366 be no formal testing of between-group differences for any variables at baseline. The main
14
15 367 comparative analyses will be carried out at a child level since outcomes relate to an
16
17 368 individual's exposure to one or more UTI and will test the null hypothesis that there are no
18
19 369 differences in outcomes in the first 5 years and between the age of 5 and 7 years. The
20
21 370 primary analysis will consist of a comparison of rate of diagnosis of renal scarring (no renal
22
23 371 scarring, renal scarring recorded < 5 years, renal scarring recorded 5-7 years) in children
24
25 372 with mcUTI versus children with no mcUTI, using a multinomial regression model. Results
26
27 373 will be reported as relative risk ratios alongside 95% confidence intervals (CIs). We will
28
29 374 adjust for direct covariates of renal scarring and explore the impact of indirect effects such
30
31 375 as a mcUTI using a causal directed acyclic graph (DAG). Unadjusted and adjusted relative
32
33 376 risk ratios will be estimated, together with 95% CIs. Cox regression will also be performed to
34
35 377 model time to first renal scarring diagnosis to allow us to look for this outcome using all
36
37 378 available follow-up (at least 7 years). We will estimate hazard ratios with 95% CIs for each
38
39 379 exposure group.

40
41
42 380 Several sensitivity analyses are proposed: The primary outcome will be expanded to include
43
44 381 any renal pathology codes due to uncertainty around whether the renal scarring codes are
45
46 382 sufficiently sensitive to pick up all cases of renal scarring. In addition we will explore using
47
48 383 Read codes from GP notes to pick up additional renal scarring diagnoses. Two effect
49
50 384 modifiers were identified as a basis for sub-group analyses for the primary outcome: gender
51
52 385 of child and presence of any renal/urological congenital anomalies. These pre-planned
53
54 386 analyses will be conducted by the inclusion of appropriate interaction terms in the models.

1
2
3 387 Secondary outcomes will be analysed using multinomial and time to event models (Table 3).
4
5 388 Poisson regression models will be used where the outcome is a count of event (e.g. hospital
6
7 389 admissions, GP consultations, antibiotics prescribed); results will be represented as
8
9 390 incidence rate ratios alongside 95% CIs. Additionally we will identify risk factors for renal
10
11 391 scarring (as above, plus age at first mcUTI, bacterial type & resistance) in children
12
13 392 presenting with a childhood mcUTI. In a sub sample of children linked to GP data, the 14
14
15 393 days prior to the urine sample submission date will be examined to determine whether there
16
17 394 was an associated GP consultation. We will then estimate the likely rate of GP associated
18
19 395 mcUTIs and the rate of mcUTIs submitted from another setting (such as out of hours,
20
21 396 outpatients or as a result of a hospital admission). An identical analysis to the primary
22
23 397 outcome will be taken to examine whether the risk of renal scarring differs between those
24
25 398 with a mcUTI or not, between different settings and the interaction between the two. Where
26
27 399 numbers allow, variation in outcome will be accounted for at the level of the general practice
28
29 400 in a multilevel model. We will also describe the levels of urine sampling and incidence of
30
31 401 mcUTI from General Practice in the routine data.

32 33 402 *Dataset 2: Systematic sampling of UTI*

34
35
36 403 Detailed study data for DUTY and EURICA participants is available including age, gender
37
38 404 and deprivation, presenting features, GP diagnosis and acute management. Recruited
39
40 405 children are already grouped into mcUTI or no mcUTI through the study microbiology data
41
42 406 and outcomes will be compared according to these groups. Five year in-patient hospital
43
44 407 outcomes will be available for all EURICA and DUTY children (the last participant of DUTY
45
46 408 study was recruited in April 2012). We will be able to describe serious short-term (less than 1
47
48 409 year) and medium-term (1-5 years) outcomes, including hospital admission, renal imaging,
49
50 410 renal scarring, VUR and renal failure outcomes for all children in this group using PEDW
51
52 411 data in Wales and NHS Digital hospital data in England. We will be able to describe other
53
54 412 outcomes, such as GP consultations, recurrent UTI, dysfunctional voiding, antibiotic
55
56 413 prescriptions, hypertension, chronic kidney disease, using GP data on a sub-set. In Wales,

1
2
3 414 Datastore will also be used to look at the urine culture results and organism resistance
4
5 415 profile for subsequent UTIs.
6

7 416 We will describe GP diagnosis from study data versus Read codes and acute management
8
9 417 from the routine data in GP records for this cohort for later comparisons and also to explore
10
11 418 the validity of using routinely collected data in these cases. We will also assess the validity of
12
13 419 using Read codes to diagnose UTI against microbiological culture results and agreement will
14
15 420 be measured using the Kappa statistic.
16

17
18 421

19
20 422 Research question 2: Comparison of short- and medium-term outcomes in children with
21
22 423 mcUTI: routine versus systematic sampling.
23

24
25 424 We will compare the outcomes in children with mcUTI identified through routine versus
26
27 425 systematic sampling. Children's characteristics, presentation factors, acute management
28
29 426 and microbiology results will be described for the groups using appropriate summary
30
31 427 statistics. In addition, we will describe blood pressure and creatinine levels for each group if
32
33 428 recorded and explore whether comparisons can be made.
34

35
36 429 Previously mentioned short- and medium-term outcomes will be described by the two groups
37
38 430 of routine vs. selective sampling. Predictors of outcome will be examined as before using a
39
40 431 multilevel multinomial regression model (no event, event <1 year, event 1-5 years) and again
41
42 432 where numbers allowed, variation in outcome will be accounted for at the level of the general
43
44 433 practice. Associations between covariates previously described and outcome will firstly be
45
46 434 examined. Unadjusted and adjusted relative risk ratios will be estimated, together with 95%
47
48 435 CIs. We will compare blood pressure and creatinine levels (where available) across the
49
50 436 groups; we expect this data to be limited so will be exploratory.
51

52 437 A detailed statistical analysis plan will be written prior to database lock. The reporting and
53
54 438 presentation of results will be in accordance with the [24–26] statements to ensure the
55
56
57
58
59

1
2
3 439 comprehensive reporting of our observational non-randomized evaluation of a public health
4 440 intervention. SPSS and Stata will be used for all analyses [27,28].
5
6
7

8 441 **ETHICS AND DISSEMINATION**

9
10 442 The governance surrounding dataset one differs from dataset two. Dataset one is an
11 443 anonymised dataset made available from SAIL databank with only approval required from
12 444 the IGRP whereas dataset two involves the transfer of identifiable data to data providers
13 445 which requires ethical approval, section 251 (s251) support, IGRP and IGARD approval. In
14 446 order to obtain an unbiased sample from the DUTY and EURICA cohort an opt-out consent
15 447 model is being used, this was supported by both the ethics panel and the CAG panel as
16 448 justification for this model of consent.
17
18
19
20
21
22
23

24 449 The LUCI study will report the risk of renal scarring for children with and without childhood
25 450 mcUTI across the whole of Wales. The linking of routinely collected datasets will give us a
26 451 large cohort including demographic, hospital in-patient and out-patient, GP and microbiology
27 452 data, allowing us to define mcUTI cases and describe outcomes for all children from both
28 453 primary and secondary care. Clarifying the link between UTI, renal scarring and long-term
29 454 complications will inform the management of acutely ill children in primary care, where the
30 455 need for urine sampling is unclear. Determining the clinical implications of 'missed' cases of
31 456 UTI through our comparison of children with mcUTI identified through routine and systematic
32 457 urine sampling will also help to determine the most appropriate urine sampling strategy.
33 458 This study maximises the benefits of the previously funded DUTY and EURICA cohorts,
34 459 representing over 8000 acutely ill children recruited from UK primary care. Significant
35 460 resources were invested by funders, patients and staff to develop these cohorts. Routine
36 461 data linkage will allow us to determine longer-term outcomes for these children and to
37 462 determine risks of adverse outcomes. It will also pave the way for even longer-term follow-up
38 463 of cohorts of children with UTI (diagnosed both systematically and routinely) which has been
39 464 identified as a high research priority by NICE.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 465 A lay summary of the results and links to publications will be made available on the
4
5 466 University project website and the participant. The academic outputs for this study include (i)
6
7 467 this protocol paper, (ii) main results from research question one and (ii) main results from
8
9 468 research question two. The findings from this study will be of interest to clinicians and policy
10
11 469 makers and may influence the management of acutely ill children and childhood UTI.
12

13 **470 DECLARATIONS**

14
15 471 *List of abbreviations*

16	ALF: Anonymised linking field
17	ALF-E: Anonymised linking field encryption
18	ALF-E: Anonymised linking field encryption
19	ALF-E: Anonymised linking field encryption
20	CIs: Confidence intervals
21	CIs: Confidence intervals
22	DAG: Directed acyclic graph
23	DAG: Directed acyclic graph
24	DOB: Date of Birth
25	DOB: Date of Birth
26	ESRF: End-stage renal failure
27	ESRF: End-stage renal failure
28	HES: Hospital Episode Statistics
29	HES: Hospital Episode Statistics
30	HRA CAG: Health Research Authority's Confidentiality Advisory Group
31	HRA CAG: Health Research Authority's Confidentiality Advisory Group
32	HSCIC: Health and Social Care Information Centre
33	HSCIC: Health and Social Care Information Centre
34	IGARD: Independent Group Advising on the Release of Data
35	IGARD: Independent Group Advising on the Release of Data
36	IGRP: Information Governance Review Panel
37	IGRP: Information Governance Review Panel
38	LSOA: Lower super output area
39	LSOA: Lower super output area
40	mcUTI: Microbiological culture urinary tract infection
41	mcUTI: Microbiological culture urinary tract infection
42	NICE: National Institute for Health and Clinical Excellence
43	NICE: National Institute for Health and Clinical Excellence
44	NIHR HTA: National Institute of Health Research Health Technology Assessment
45	NIHR HTA: National Institute of Health Research Health Technology Assessment
46	NISCHR: National Institute for Social Care and Health Research
47	NISCHR: National Institute for Social Care and Health Research
48	NWIS: NHS Wales Informatics Service
49	NWIS: NHS Wales Informatics Service
50	PEDW: Patient Episode Database for Wales
51	PEDW: Patient Episode Database for Wales
52	PRIME: Primary and Emergency Care Research
53	PRIME: Primary and Emergency Care Research
54	SAIL: Secure Anonymised Information Linkage
55	SAIL: Secure Anonymised Information Linkage
56	TTP: Trusted third party
57	TTP: Trusted third party
58	UTI: Urinary tract infection
59	UTI: Urinary tract infection
60	VUR: Vesicoureteric reflux

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

472

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

477 *Consent for publication* - Not Applicable

478 *Availability of data and material* - Not Applicable

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

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

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

487 **Acknowledgements**

488

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

1
2
3 495 National Institute of Health Research, the National Institute for Social Care and Health
4
5 496 Research (Welsh Assembly Government), the Chief Scientist Office (Scottish Government
6
7 497 Health Directorates), and the Wellcome Trust, (MRC Grant No: MR/K006525/1) and the
8
9 498 National Centre for Population Health & Wellbeing Research (NCPHWR).

10
11 499

12
13 **References**

- 14
15 501 1 Craig JC, Williams GJ, Jones M, *et al.* The accuracy of clinical symptoms and signs
16
17 502 for the diagnosis of serious bacterial infection in young febrile children: prospective
18
19 503 cohort study of 15 781 febrile illnesses. *BMJ* 2010;**340**:c1594.
20
21 504 doi:10.1136/BMJ.C1594
- 22 505 2 Hoberman A, Chao HP, Keller DM, *et al.* Prevalence of urinary tract infection in febrile
23
24 506 infants. *J Pediatr* 1993;**123**:17–23.<http://www.ncbi.nlm.nih.gov/pubmed/8320616>
25
26 507 (accessed 17 Apr 2018).
- 27 508 3 Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and
28
29 509 economic costs. *Am J Med* 2002;**113 Suppl 1A**:5S–
30
31 510 13S.<http://www.ncbi.nlm.nih.gov/pubmed/12113866> (accessed 17 Apr 2018).
- 32 511 4 Hay AD, Birnie K, Busby J, *et al.* The Diagnosis of Urinary Tract infection in Young
33
34 512 children (DUTY): a diagnostic prospective observational study to derive and validate a
35
36 513 clinical algorithm for the diagnosis of urinary tract infection in children presenting to
37
38 514 primary care with an acute illness. *Health Technol Assess (Rockv)* 2016;**20**:1–294.
39
40 515 doi:10.3310/hta20510
- 41 516 5 O'Brien K, Edwards A, Hood K, *et al.* Prevalence of urinary tract infection in acutely
42
43 517 unwell children in general practice: a prospective study with systematic urine
44
45 518 sampling. *Br J Gen Pract* 2013;**63**:e156–64. doi:10.3399/bjgp13X663127
- 46 519 6 Urinary tract infection in under 16s: diagnosis and management | Guidance and
47
48 520 guidelines | NICE. <https://www.nice.org.uk/guidance/cg54> (accessed 24 Apr 2018).
- 49 521 7 Butler CC, O'Brien K, Pickles T, *et al.* Childhood urinary tract infection in primary care:
50
51 522 a prospective observational study of prevalence, diagnosis, treatment, and recovery.
52
53 523 *Br J Gen Pract* 2015;**65**:e217–23. doi:10.3399/bjgp15X684361
- 54 524 8 Shaikh N, Ewing AL, Bhatnagar S, *et al.* Risk of Renal Scarring in Children With a
55
56 525 First Urinary Tract Infection: A Systematic Review. *Pediatrics* 2010;**126**:1084–91.

- 1
2
3 526 doi:10.1542/peds.2010-0685
4
5 527 9 Coulthard MG, Lambert HJ, Keir MJ. Occurrence of renal scars in children after their
6 528 first referral for urinary tract infection. *BMJ* 1997;**315**:918–
7 529 9.<http://www.ncbi.nlm.nih.gov/pubmed/9361542> (accessed 17 Apr 2018).
8
9
10 530 10 Jacobson SH, Eklöf O, Eriksson CG, *et al.* Development of hypertension and uraemia
11 531 after pyelonephritis in childhood: 27 year follow up. *BMJ* 1989;**299**:703–
12 532 6.<http://www.ncbi.nlm.nih.gov/pubmed/2508881> (accessed 17 Apr 2018).
13
14
15 533 11 Round J, Fitzgerald AC, Hulme C, *et al.* Urinary tract infections in children and the risk
16 534 of ESRF. *Acta Paediatr* 2012;**101**:278–82. doi:10.1111/j.1651-2227.2011.02542.x
17
18
19 535 12 Smellie JM, Prescod NP, Shaw PJ, *et al.* Childhood reflux and urinary infection: a
20 536 follow-up of 10–41 years in 226 adults. *Pediatr Nephrol* 1998;**12**:727–
21 537 36.<http://www.ncbi.nlm.nih.gov/pubmed/9874316> (accessed 17 Apr 2018).
22
23
24 538 13 Craig JC, Williams GJ. Denominators do matter: it's a myth—urinary tract infection
25 539 does not cause chronic kidney disease. *Pediatrics* 2011;**128**:984–5.
26 540 doi:10.1542/peds.2011-2631
27
28
29 541 14 Salo J, Ikäheimo R, Tapiainen T, *et al.* Childhood urinary tract infections as a cause of
30 542 chronic kidney disease. *Pediatrics* 2011;**128**:840–7. doi:10.1542/peds.2010-3520
31
32
33 543 15 Coulthard MG, Vernon SJ, Lambert HJ, *et al.* A nurse led education and direct access
34 544 service for the management of urinary tract infections in children: prospective
35 545 controlled trial. *BMJ* 2003;**327**:656. doi:10.1136/bmj.327.7416.656
36
37
38 546 16 Jadresic L, Cartwright K, Cowie N, *et al.* Investigation of urinary tract infection in
39 547 childhood. *BMJ* 1993;**307**:761–4. doi:10.1136/BMJ.307.6907.761
40
41
42 548 17 van der Voort J, Edwards A, Roberts R, *et al.* The struggle to diagnose UTI in children
43 549 under two in primary care. *Fam Pract* 1997;**14**:44–
44 550 8.<http://www.ncbi.nlm.nih.gov/pubmed/9061344> (accessed 17 Apr 2018).
45
46
47 551 18 Hollingworth W, Busby J, Butler CC, *et al.* The Diagnosis of Urinary Tract Infection in
48 552 Young Children (DUTY) Study Clinical Rule: Economic Evaluation. *Value Heal*
49 553 2017;**20**:556–66. doi:10.1016/j.jval.2017.01.003
50
51
52 554 19 Jones KH, Ford D V., Jones C, *et al.* A case study of the Secure Anonymous
53 555 Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for
54 556 health-related research and evaluation. *J Biomed Inform* 2014;**50**:196–204.
55 557 doi:10.1016/j.jbi.2014.01.003
56
57
58
59
60

- 1
2
3 558 20 Ford D V, Jones KH, Verplancke J-P, *et al.* The SAIL Databank: building a national
4 559 architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;**9**:157.
5 560 doi:10.1186/1472-6963-9-157
6
7
8 561 21 Lyons RA, Jones KH, John G, *et al.* The SAIL databank: linking multiple health and
9 562 social care datasets. *BMC Med Inform Decis Mak* 2009;**9**:3. doi:10.1186/1472-6947-
10 563 9-3
11
12
13 564 22 Robling M, Bekkers M-J, Bell K, *et al.* Effectiveness of a nurse-led intensive home-
14 565 visitation programme for first-time teenage mothers (Building Blocks): a pragmatic
15 566 randomised controlled trial. *Lancet* 2016;**387**:146–55. doi:10.1016/S0140-
16 567 6736(15)00392-X
17
18
19 568 23 Green SB. How Many Subjects Does It Take To Do A Regression Analysis.
20 569 *Multivariate Behav Res* 1991;**26**:499–510. doi:10.1207/s15327906mbr2603_7
21
22
23 570 24 Gilbert R, Lafferty R, Hagger-Johnson G, *et al.* GUILD: GUIDance for Information
24 571 about Linking Data sets†. *J Public Health (Bangkok)* 2018;**40**:191–8.
25 572 doi:10.1093/pubmed/fox037
26
27
28 573 25 von Elm E, Altman DG, Egger M, *et al.* Strengthening the Reporting of Observational
29 574 Studies in Epidemiology (STROBE) statement: guidelines for reporting observational
30 575 studies. *BMJ* 2007;**335**:806–8. doi:10.1136/bmj.39335.541782.AD
31
32
33 576 26 Benchimol EI, Smeeth L, Guttman A, *et al.* The REporting of studies Conducted
34 577 using Observational Routinely-collected health Data (RECORD) Statement. *PLOS*
35 578 *Med* 2015;**12**:e1001885. doi:10.1371/journal.pmed.1001885
36
37
38 579 27 IBM. IBM SPSS Statistics for Windows: Version 22.0. 2013.
39
40 580 28 StataCorp. Stata Statistical Software: Release 15. 2017. 2017. doi:10.2307/2234838
41
42

43 **582**
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 583 *(Title)* Figure 1. The data flow for dataset 2.
4

5 584 *(Legend)* ALF- Anonymised Linking Field; NWIS – NHS Wales Informatics Service; HES –
6
7 585 Hospital Episode Statistics; SAIL – Secure Anonymised Information Linkage
8
9

10 586
11

12 587 *(Title)* Figure 2. Flow chart of study participants
13
14

15 588
16

17 589
18

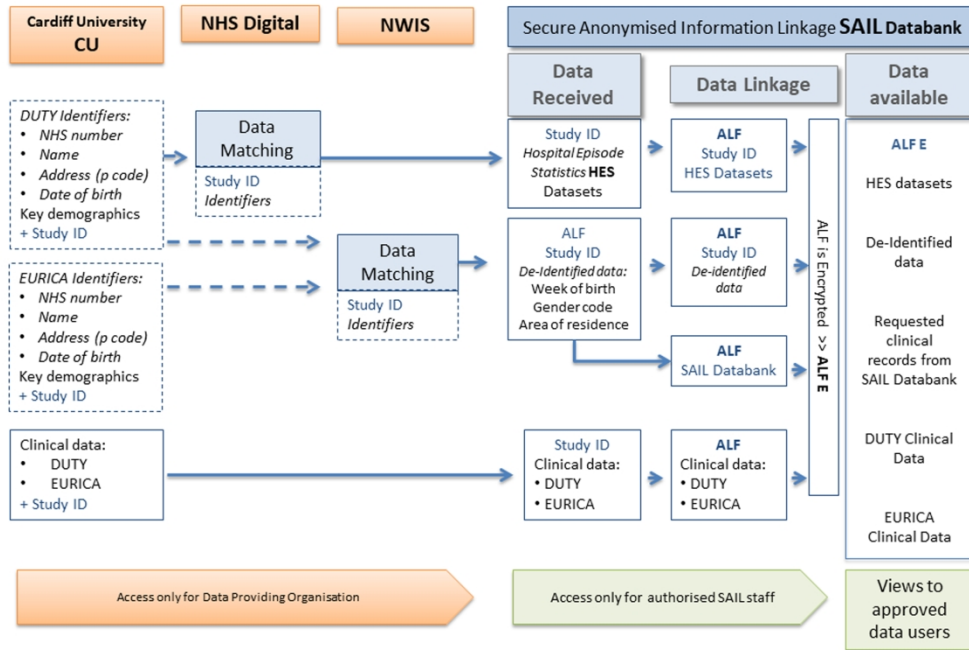
19 590
20

21 591
22

22 592
23

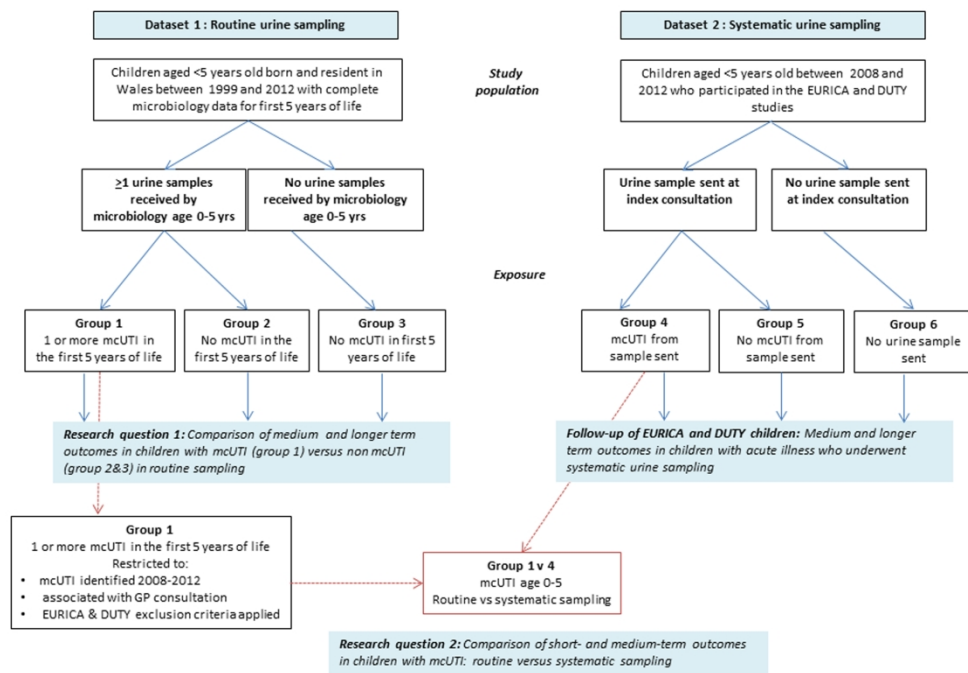
24 593
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only



(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:	<i>BMJ Open</i>
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 Heginbotham, 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

SCHOLARONE™
Manuscripts

1
2
3 1 Long-term outcomes of urinary tract infection (UTI) in Childhood (LUCI): protocol for an
4 2 electronic record-linked cohort study
5
6 3
7

8 4 **Corresponding author:**

9 5 Fiona Lugg-Widger LuggFV@cardiff.ac.uk

10 6 Address: 4th Floor Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4YS.

11
12 7
13 8 Lianna Angel¹ AngelL@cardiff.ac.uk
14 9 Rebecca Cannings-John¹ CanningsRL@cardiff.ac.uk
15 10 Hywel Jones² JonesH75@cardiff.ac.uk
16 11 Mandy Lau¹ LauTM@cardiff.ac.uk
17 12 Christopher C Butler³ christopher.butler@phc.ox.ac.uk
18 13 Nick Francis⁴ FrancisNA@cardiff.ac.uk
19 14 Alastair D Hay⁵ alastair.hay@bristol.ac.uk
20 15 Margaret Heginbothom⁶ margaret.heginbothom@wales.nhs.uk
21 16 Kerenza Hood¹ HoodK1@cardiff.ac.uk
22 17 Shantini Paranjothy² ParanjothyS@cardiff.ac.uk
23 18 Judith Van der Voort⁷ judith.vandervoort@wales.nhs.uk
24 19 Kathryn Hughes⁴ HughesKA6@cardiff.ac.uk
25
26
27

28
29 21 **Author Affiliations**

- 30 22 1. Centre for Trials Research, College of Biomedical & Life Sciences, Cardiff University,
31 23 Cardiff, UK.
32 24 2. National Centre for Population Health and Wellbeing Research, Division of
33 25 Population Medicine, Cardiff University
34 26 3. Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe
35 27 Primary Care Building, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2
36 28 6GG
37 29 4. Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales),
38 30 Division of Population Medicine, School of Medicine, Cardiff University
39 31 5. Centre of Academic Primary Care, NIHR School for Primary Care Research, Bristol
40 32 Medical School: Population Health Sciences, Canynge Hall, 39 Whatley Road, Bristol
41 33 BS8 2PS
42 34 6. Public Health Wales, 2 Capital Quarter, Tyndall Street, Cardiff CF10 4BZ
43 35 7. Noah's Arc Childrens Hospital for Wales, Cardiff, CF14 4XW
44
45
46
47
48
49
50
51
52

53 39 **Word Count:** 6810

54 40 **Keywords:** Children; Urinary Tract Infection; Medical record linkage; Renal scarring; Urine
55 41 sampling; Primary care.
56
57
58

59 43 **ABSTRACT**
60

1
2
3 44 **Introduction:** Current guidelines advise the prompt diagnosis and treatment of UTI in
4
5 45 children to improve both short and longer term outcomes. However, the risk of long-term
6
7 46 complications following childhood UTI is unclear.

8
9
10 47 UTI is relatively common but difficult to diagnose in children as symptoms are non-specific.
11
12 48 Diagnosis requires a urine sample, but sampling is difficult and infrequent, and it is not clear
13
14 49 if sampling should be given greater priority in primary care. The LUCI study will assess the
15
16 50 short, medium and longer-term outcomes of childhood UTI associated with routine and
17
18 51 systematic sampling practices.

19
20
21 52 **Methods and analysis:** Two datasets will be established: The first will consist of routinely
22
23 53 collected data (Hospital, GP, Microbiology) from children born and resident in Wales, linked
24
25 54 via the Secure Anonymised Information Linkage (SAIL) databank (an 'e-cohort'). Urine
26
27 55 sampling in this dataset reflects normal practice 'routine sampling'. Outcomes (including
28
29 56 renal scarring, hypertension, end-stage renal failure(ESRF), hospital admissions, GP
30
31 57 consultations, antibiotic prescriptions) for children with at least one UTI confirmed with
32
33 58 microbiological culture (mcUTI) or no mcUTI before the age of 5 will be compared.

34
35
36
37 59 The second will combine data from two prospective observational studies('DUTY' &
38
39 60 'EURICA') employing systematic urine sampling for children presenting to primary care with
40
41 61 acute, undifferentiated illness, linked to routine data via SAIL (Wales) and NHS Digital
42
43 62 (England). Outcomes (as above, plus features of mcUTI) for children with a mcUTI in this
44
45 63 dataset, identified through systematic urine sampling, will be compared to those with a
46
47 64 mcUTI identified through routine urine sampling (dataset 1).

48
49
50 65 **Ethics and dissemination:** The study protocol has been approved by NHS Wales
51
52 66 Research Ethics Committee and the Health Research Authority's Confidentiality Advisory
53
54 67 Group. Methods of innovative study design and findings will be disseminated through peer-
55
56 68 review journals and conferences. Results will be of interest to clinical and policy
57
58 69 stakeholders in the UK.

70 **Strengths and limitations of this study:**

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

85 **INTRODUCTION**

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

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

1
2
3 96 were conducted in secondary care and most required fever for inclusion in the study [8].
4
5 97 These are likely to represent more serious UTIs than those presenting in primary care, and
6
7 98 this rate of renal scarring may not apply to all children presenting with UTI. The risk of long-
8
9 99 term complications (such as hypertension and renal failure) as a result of renal scarring
10
11 100 following childhood UTI is also not clear, with some researchers questioning this association.
12
13 101 [6,13,14] A review, commissioned by the National Institute for Health and Clinical Excellence
14
15 102 (NICE), concluded that 'there are no appropriate studies that accurately estimate the risks of
16
17 103 long term complications as a result of childhood UTI'. [6]

18
19
20 104 A urine sample is necessary to confirm the presence of UTI in childhood as the presenting
21
22 105 symptoms are non-specific and similar to those found in many common childhood illnesses
23
24 106 (e.g. respiratory infections, gastroenteritis, tonsillitis, ear infections) such as high
25
26 107 temperature, poor feeding, vomiting, lethargy and irritability. [6] Furthermore, microbiological
27
28 108 confirmation is important as some children with UTI require invasive investigations. [6]
29
30 109 Children who become seriously ill and who are assessed in emergency departments or
31
32 110 admitted to hospital will usually have their urine sampled. [1] However, in primary care,
33
34 111 where most acute illness is seen, urine is infrequently sampled. It is estimated that urine is
35
36 112 sampled in fewer than 2% of acute illness consultations with children under five years old in
37
38 113 the UK. [4] Studies have suggested that many cases of childhood UTI are missed in primary
39
40 114 care. [15–17] Routine practice (urine sampling based on clinician suspicion) is likely to miss
41
42 115 more than two thirds of UTIs in primary care. [5,7] The clinical implications of missed
43
44 116 childhood UTIs are not known. Increasing urine sampling in primary care is likely to increase
45
46 117 the diagnosis of childhood mcUTI, and is advocated by current guidelines, but it is not clear
47
48 118 whether this is an appropriate strategy, whether outcomes for children would improve or to
49
50 119 what extent it should be increased. [4–6,15,18]

51
52
53
54
55 120 Clarification is needed on two issues: First, to determine what the longer-term outcomes are
56
57 121 following childhood UTI (including those identified in primary care), and second, to determine
58
59 122 whether outcomes vary according to sampling practice.
60

1
2
3 123 In this study we will describe clinical outcomes for all children with one or more mcUTI aged
4
5 124 less than five years old, compared to those with no mcUTI, using NHS laboratory data from
6
7 125 across Wales. We will examine the risk factors for being diagnosed with renal scarring
8
9 126 following mcUTI.

11
12 127 We will also describe longer-term follow up of clinical outcomes for at least five years
13
14 128 following participation in two UK prospective cohort studies of acutely ill children with
15
16 129 systematic urine sampling in primary care, the DUTY and EURICA studies. [5,7] We will
17
18 130 compare the outcomes of those with mcUTI identified through these studies (systematic
19
20 131 urine sampling) with those identified through routine practice.

23 132 **METHODS AND DESIGN**

24 133 **Research objectives**

25
26
27
28
29 134 The LUCI Study will use data linkage of routinely collected datasets and data from two
30
31 135 cohorts of participants to answer two main research questions:

32
33
34 136 Research Question 1: Through routine sampling, do children who have experienced a
35
36 137 mcUTI aged <5 years old have worse outcomes (medium (0-5 years); longer-term (>5
37
38 138 years)) compared to children who have not experienced a mcUTI?

39
40
41 139 Research Question 2: Are short-term (<12 months) and medium-term (1-4 years) outcomes
42
43 140 different for children with childhood mcUTI identified through systematic sampling compared
44
45 141 to routine sampling (standard, clinician-led sampling)?

46 47 48 142 **Study Design**

49
50
51 143 This is a data linkage study comprising two overarching datasets of children. Dataset 1 will
52
53 144 comprise routinely collected health data from children born and resident in Wales. Children
54
55 145 in this dataset will have had urine sampled according to routine practice. Routine data will be
56
57 146 available on all children for seven years, and longer for some (i.e. children will be followed
58
59 147 up until the date of data abstraction). Dataset 1 will be used to answer Research Question 1.
60

1
2
3 148 Dataset 2 will be children who participated in the EURICA or DUTY studies. Children in this
4
5 149 dataset had their urine systematically sampled (all children presenting with an acute illness
6
7 150 were asked to provide a urine sample). DUTY and EURICA children will be followed-up by
8
9 151 linking records to routinely collected health data from England (using NHS Digital) and
10
11 152 Wales (using SAIL). Follow-up will be for a minimum of five years, and longer where data is
12
13 153 available. Dataset 2 will be used to describe longer term follow-up for DUTY and EURICA
14
15 154 study children. Those with mcUTI in Dataset 2 will be compared to those with mcUTI
16
17 155 identified in Dataset 1 to answer Research Question 2.
18
19

20 156 The study formally started in October 2016 and will report to funder in December 2018. A
21
22 157 summary of the data sources is provided in Table 1.
23
24

25 158

26 159

27 160

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

161 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

For peer review only

163 **Data providers and datasets**

164 The EURICA and DUTY Studies

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

178

179 SAIL Databank

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

188

1
2
3 189 NHS Digital
4
5

6 190 NHS Digital is the trading name of the Health and Social Care Information Centre (HSCIC).
7
8 191 This study will access Hospital Episode Statistics (HES) data for participants of the DUTY
9
10 192 and EURICA study. All available Inpatient and Outpatient records belonging to each study
11
12 193 participant will be requested and approved by the Independent Group Advising on the
13
14 194 Release of Data (IGARD) Panel. The data requested includes diagnosis, procedures and
15
16 195 length of episode according to the 10th revision of the International Statistical Classification
17
18 196 of Diseases and Related Health Problems [ICD-10] codes.
19

20
21 197
22
23

24 198 Public Health Wales
25

26
27 199 Public Health Wales will provide a data extract of urine microbiology culture results from all
28
29 200 microbiology laboratories in Wales (Datastore) for use with this project. This will be
30
31 201 transferred to SAIL.
32

33 202
34

35 203 Individual Health Boards
36

37
38 204 Health boards in Wales will be approached to access anonymised radiology data for patients
39
40 205 in dataset 1. A one off data extract of patient-level attendance data for patients born between
41
42 206 01/01/1994 and 31/12/2012 that attended radiology between 1994 and 2016 will be
43
44 207 transferred to SAIL. Data extracted includes examination performed, attendance data and
45
46 208 the radiology report.
47

48
49 209
50

51
52 210 **Opportunity to opt-out (dataset two)**
53

54
55 211 Dataset one uses routinely collected data that is fully anonymised so we do not require
56
57 212 individual consent in order to access these data. Dataset two involves participants from the
58
59 213 DUTY and EURICA study and therefore requires section 251 (s251) support of the 2006
60

1
2
3 214 NHS Act approval from the Health Research Authority's Confidentiality Advisory Group (HRA
4
5 215 CAG) to pass identifiable participant data legally held by Cardiff and Bristol University (as
6
7 216 sponsors of the studies) to the data providers. This is using an opt-out/dissent model instead
8
9 217 of obtaining further consent. In order to provide the opportunity for participants to dissent a
10
11 218 letter has been developed to be sent to all parents of participants. Participants have the
12
13 219 opportunity to contact the study team through an online web-form, by email, text or
14
15 220 telephone (details provided on website) and register their dissent. These participants who
16
17 221 register their dissent will not be included in any of the datasets sent to the information
18
19 222 centres and therefore will not appear in the datasets for analysis. A participant
20
21 223 representative was consulted on the layout, wording and level of information contained in the
22
23 224 participant opt-out letter and on the study website. A key consideration was to communicate
24
25 225 the data transfer process. The final letter was approved by both an NHS Research Ethics
26
27 226 Committee and CAG committee as part of overall governance approval for the study.
28
29
30

31
32
33

34 228 **Data Matching**

35
36
37 229 For included data sets, data will be sent to NHS Digital and SAIL (via NWIS their trusted
38
39 230 third party (TTP)) for matching using a combination of NHS number, name, address and
40
41 231 date of birth. Matching with NHS Digital data will be by exact matching on NHS Number,
42
43 232 Date of Birth and Postcode. This has been conducted in other studies and achieved a high
44
45 233 match rate (99.6%) [22]. NWIS match using NHS Number, Date of Birth, Sex, Postcode.
46
47 234 Data matching by NHS Digital will be separate to matching conducted by NWIS on behalf of
48
49 235 SAIL. We would expect only a small number of participants matched to both English and
50
51 236 Welsh NHS records however there is the possibility of this for those using services across
52
53 237 the border to their current address.
54
55

56
57
58

59 239 **The anonymised dataset**

1
2
3 240 For DUTY and EURICA participants NHS Digital will retain the unique study ID assigned by
4
5 241 the DUTY/EURICA study, with the HES records after the matching process before it is sent
6
7 242 to SAIL Databank. This ID will be retained for data transfer to ensure the data can be linked.
8
9 243 The same applies to data sent to NWIS. Requested data will then be linked for each
10
11 244 participant via SAIL using the unique study ID. The DUTY and EURICA clinical datasets will
12
13 245 be transferred to SAIL following a process of de-identification. The data flow is shown in
14
15 246 Figure 1. The study ID will be replaced by an anonymised linking field (ALF) to enable all
16
17 247 datasets to be linked at an individual level. The resulting dataset will contain clinical
18
19 248 variables (from DUTY, EURICA, SAIL and NHS Digital), de-identified demographic variables
20
21 249 (e.g. week of birth and lower super output area - LSOA) and the ALF. The key between
22
23 250 study ID and ALF will be retained and encrypted as a further safeguard. The ALF is further
24
25 251 encrypted (ALF-E) so that datasets and records cannot be linked across multiple projects a
26
27 252 researcher has access to.
28
29
30
31 253

32 33 254 **Study Participants**

34
35
36 255 A flow chart of participants in the study is shown in Figure 2. Dataset 1 (routine sampling)
37
38 256 will be children born and resident in Wales for the first 5 years of life; who were less than 5
39
40 257 years old between 1999 and 2012. The main analysis will be on children born between
41
42 258 01/01/2005 and 31/12/2009 to ensure that all of the first five years of life are covered by the
43
44 259 dates when Datastore is available.

45
46
47 260 Dataset 2 (systematic sampling from DUTY and EURICA) will be children who participated in
48
49 261 the EURICA or DUTY studies, who were not withdrawn from the study and when provided
50
51 262 with the opportunity to opt-out, remained in the study. For research question 2 when
52
53 263 comparing children with a mcUTI from Dataset 1 and 2, children in Group 1 will be selected
54
55 264 to match the DUTY and EURICA study eligibility criteria as closely as possible within the
56
57 265 constraints of using routine data. Datasets 1 and 2 will be limited to children with a mcUTI
58
59 266 associated with a GP consultation between March 2008 and April 2012. For both datasets,

1
2
3 267 the child's mcUTI, associated with a GP consultation (defined as within 14 days prior to the
4
5 268 sample) will be identified and defined as the index consultation. To limit the potential
6
7 269 transfer of GP systematic sampling behaviour, children from Group 1 with index
8
9 270 consultations between 2008 and 2012 at practices which participated in the EURICA study
10
11 271 will be flagged as will children with index consultations between 2010 and 2012 at practices
12
13 272 which participated in the DUTY study. Children will only be included once in each study
14
15 273 period (i.e. a child with a sample sent within the EURICA study period could also appear in
16
17 274 the DUTY study period). In addition, we will apply the DUTY study exclusion criteria (where
18
19 275 possible) to Dataset 1 to make them directly comparable with Dataset 2; therefore children
20
21 276 will be excluded from the routine sampling cohort if any criterion met:
22
23
24

277

- 27 278 • Known neurogenic bladder (e.g. spina bifida) or previous bladder surgery
- 28 279 • Prescribed antibiotics in the 7 days prior to presentation
- 29 280 • Taking immunosuppressant medication
- 30 281 • Using urinary catheters

282

31
32
33
34
35
36
37 283 Children with a prior history of UTI or vesicoureteric reflux (VUR) will not be excluded (and
38
39 284 were not excluded from the DUTY or EURICA studies) and we will explore the impact of
40
41 285 these risk factors on outcomes.
42
43

286

287

288

289 **Exposure**

44
45
46
47
48
49
50 290 UTI cases will be based on NHS laboratory results. For Dataset 1, this will be through
51
52 291 microbiological culture data downloaded from Datastore. These data represent samples
53
54 292 (from both community and hospital settings) which have been classified as positive or
55
56 293 negative by NHS laboratories according to their standard operating procedures. We do not
57
58 294 know how urine was sampled, and this is likely to vary between settings. In most cases,
59
60

1
2
3 295 these are likely to be clean catch samples, but may include urine collection pads or bags
4
5 296 (particularly in community samples) as recommended by NICE; or catheter or suprapubic
6
7 297 aspiration (SPA) in hospital samples.[6] NHS laboratories take into consideration the nature
8
9 298 of the urine sample in their reporting.[23] For Dataset 2, we will use the results of
10
11 299 microbiological culture from NHS laboratories collected during the DUTY and EURICA
12
13 300 studies as some participants were from England (Datastore is Wales only). For Dataset 2,
14
15 301 the presence of significant bacteriuria (pure or predominant growth of 100,000 cfu/ml of
16
17 302 urine) will be used to define UTI.

18
19
20
21 303 For Dataset 1 we define the exposure period as <5 years and will be grouped as follows:
22
23 304 (Figure 1).

24
25 305 Group 1: children with at least one mcUTI before their 5th birthday or before outcome
26
27 306 of interest

28
29
30 307 Group 2: children with at least one urine sample but no mcUTI before their 5th
31
32 308 birthday or before outcome of interest

33
34
35 309 Group 3: children with no urine samples before their 5th birthday or before outcome of
36
37 310 interest

38
39
40 311 Exposure is a discrete time-varying covariate (0 (group 2 or 3 no UTI or no sample
41
42 312 respectively) until first exposure of group 1 (mcUTI) thereafter) and each patient's exposure
43
44 313 status will be taken at the point of each outcome; otherwise the exposure status of the child
45
46 314 at their 5th birthday will be taken.

47
48
49 315 For the main analyses, Groups 2 and 3 will be considered together as having no
50
51 316 microbiologically confirmed UTI. Dataset 2 will similarly be divided into three groups based
52
53 317 on their index consultation when recruited into the DUTY and EURICA studies:

54
55
56 318 Group 4: children with a mcUTI

57
58
59 319 Group 5: children who had a urine sample but no mcUTI
60

320 Group 6: children who had no urine sample

321

322

323 **Study variables**

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

328

329

330

331 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)		
<i>known to be associated</i> 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
<i>possibly associated</i> 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*
<i>Factors for follow-up of study participants & Research Question 2</i>		
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	✓		
<i>Sensitivity analyses</i>			
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

1
2
3 336 Follow-up of all children in Datasets 1 and 2 will continue until the first occurrence of any of:
4
5 337 outcome, migration, death or end of follow-up; and for the sub-analysis of GP data, if the
6
7 338 patient leaves the GP practice linked to SAIL or the last data collection from the general
8
9 339 practice. For the analysis of Research Question 1, using only children whose whole first five
10
11 340 years of life were covered by the dates that Datastore was available (excluding children
12
13 341 whos first five years of life fall outside of Datastore availability) and limiting DOBs to 1/1/2005
14
15 342 to 31/12/2009, will mean that shortest follow-up period will be 7 years. Follow-up will be
16
17 343 longer where data is available. For Research Question 2, we will examine outcomes at 30
18
19 344 days and at 1 year (short term), 1-5 years (medium-term) and >5 years (long-term) after the
20
21 345 index consultation.
22
23
24
25

26 347 **Analysis**

27 28 348 Sample size

29 30 349 *Comparison of renal scarring in children routinely sampled for UTI with and without mcUTI:*

31 350 The sample size is based on the outcome of renal scarring of children with and without
32
33 351 mcUTI and taken as 15% [8]. Using an odds ratio of 1.5 (a conservative effect size of 5%
34
35 352 difference between groups), 2-sided alpha 0.05, power=90%, and given a ratio of 32:1
36
37 353 (children diagnosed with mc UTI: children with no UTI diagnosis) in the dataset, we require
38
39 354 15,519 children (452 with mcUTI:15,067 without mcUTI) for analysis. Initial examination of
40
41 355 the SAIL dataset identified just under 13,000 children less than five years old with UTI
42
43 356 between 1999 and 2012. As our sample size calculation requires 452 with UTI, we can be
44
45 357 confident of adequate power for this study. However, the true proportion with renal scarring
46
47 358 is likely to be less than that reported by Shaikh [8]. Assuming a proportion of 8% with renal
48
49 359 scarring in the mcUTI group, and an odds ratio of 1.5 (2.5% difference), 2-sided alpha=0.05,
50
51 360 power=90%, and a ratio of 32:1, 35,833 children (1,044 with UTI: 34,789 without UTI) are
52
53 361 required for analysis, which is still achievable.
54
55
56
57
58
59
60

1
2
3 362 *Comparison of systematically versus routinely sampled UTI:* This sample size is constrained
4
5 363 by the number of children with a systematically sampled microbiologically confirmed UTI by
6
7 364 NHS lab (N=374). Using the lower estimate of renal scarring of 8% of routinely sampled
8
9 365 UTIs, then a 5% difference (13% in the systematically sampled group) would give 89%
10
11 366 power, with a 2-sided alpha=0.05 and a ratio of 32:1 (routinely sampled UTIs 11,968:
12
13 367 systematically sampled UTIs 374).

14
15
16 368 Both analyses use multivariable regression. Using Green's [24] formulae, assuming medium
17
18 369 effect sizes (odds ratio=1.5), a 2-sided alpha=0.05, and power=80%, to test 20 predictors in
19
20 370 the multivariable regression model we require at least 159 children in total. This suggests
21
22 371 that we will be adequately powered for both analyses (given these assumptions) to examine
23
24 372 predictors of short and medium-term outcomes.

25
26
27
28 373

29
30 374 Statistical analysis

31
32
33 375 *Dataset 1: Routine sampling of UTI*

34
35
36 376 Research question 1: Comparison of medium (up to 5 years) and longer term (≥ 5 years)
37
38 377 outcomes in children with mcUTI versus those without mcUTI in routine sampling.

39
40
41 378 Baseline variables will be described using appropriate descriptive summaries (N (%), mean
42
43 379 (SD), median (interquartile range)) to summarise the population for the main analyses by
44
45 380 group 1, 2 and 3 and any marked imbalance between the groups will be identified. There will
46
47 381 be no formal testing of between-group differences for any variables at baseline. The main
48
49 382 comparative analyses will be carried out at a child level since outcomes relate to an
50
51 383 individual's exposure to one or more UTI and will test the null hypothesis that there are no
52
53 384 differences in outcomes in the first 5 years and between the age of 5 and 7 years. The
54
55 385 primary analysis will consist of a comparison of rate of diagnosis of renal scarring (no renal
56
57 386 scarring, renal scarring recorded <5 years, renal scarring recorded 5-7 years) in children
58
59
60

1
2
3 387 with mcUTI versus children with no mcUTI, using a multinomial regression model. Results
4
5 388 will be reported as relative risk ratios alongside 95% confidence intervals (CIs).
6
7

8 389 We will construct a Direct Acyclic Graph (DAG) to inform our choice of variables to include in
9
10 390 the analysis. Confounding variables such as those listed in table 2 and also mcUTI that
11
12 391 could be considered to be on the causal pathway will be defined a priori.
13
14

15 392 We will run multiple mediation analyses using renal scarring as the dependent variable,
16
17 393 mcUTI as the mediation variable and confounders as the independent variables. First we will
18
19 394 identify the independent variables associated with renal scarring (using an univariable
20
21 395 logistic regression and identify the mediation variables (mcUTI or not) that are associated
22
23 396 with the significant independent variables. These will all be included in the mediation model.
24
25 397 For each of the significant independent variables, two regression models will be performed
26
27 398 with and without the mediation variable. We will calculate the indirect effect (and the effect of
28
29 399 the mediator) using the logistic regression coefficients from both regression models.
30
31

32 400 Unadjusted and adjusted relative risk ratios will be estimated, together with 95% CIs. Cox
33
34 401 regression will also be performed to model time to first renal scarring diagnosis to allow us to
35
36 402 look for this outcome using all available follow-up (at least 7 years). We will estimate hazard
37
38 403 ratios with 95% CIs for each exposure group.
39
40

41 404 Several sensitivity analyses are proposed: The primary outcome will be expanded to include
42
43 405 any renal pathology codes due to uncertainty around whether the renal scarring codes are
44
45 406 sufficiently sensitive to pick up all cases of renal scarring. In addition we will explore using
46
47 407 Read codes from GP notes to pick up additional renal scarring diagnoses. Two effect
48
49 408 modifiers were identified as a basis for sub-group analyses for the primary outcome: gender
50
51 409 of child and presence of any renal/urological congenital anomalies. These pre-planned
52
53 410 analyses will be conducted by the inclusion of appropriate interaction terms in the models.
54
55

56
57 411 Secondary outcomes will be analysed using multinomial and time to event models (Table 3).
58
59 412 Poisson regression models will be used where the outcome is a count of event (e.g. hospital
60

1
2
3 413 admissions, GP consultations, antibiotics prescribed); results will be represented as
4
5 414 incidence rate ratios alongside 95% CIs. Additionally we will identify risk factors for renal
6
7 415 scarring (as above, plus age at first mcUTI, bacterial type & resistance) in children
8
9 416 presenting with a childhood mcUTI. In a sub sample of children linked to GP data, the 14
10
11 417 days prior to the urine sample submission date will be examined to determine whether there
12
13 418 was an associated GP consultation. We will then estimate the likely rate of GP associated
14
15 419 mcUTIs and the rate of mcUTIs submitted from another setting (such as out of hours,
16
17 420 outpatients or as a result of a hospital admission). An identical analysis to the primary
18
19 421 outcome will be taken to examine whether the risk of renal scarring differs between those
20
21 422 with a mcUTI or not, between different settings and the interaction between the two. Where
22
23 423 numbers allow, variation in outcome will be accounted for at the level of the general practice
24
25 424 in a multilevel model. We will also describe the levels of urine sampling and incidence of
26
27 425 mcUTI from General Practice in the routine data.

30
31 426 *Dataset 2: Systematic sampling of UTI*

32
33
34 427 Detailed study data for DUTY and EURICA participants is available including age, gender
35
36 428 and deprivation, presenting features, GP diagnosis and acute management. Recruited
37
38 429 children are already grouped into mcUTI or no mcUTI through the study microbiology data
39
40 430 and outcomes will be compared according to these groups. Five-year in-patient hospital
41
42 431 outcomes will be available for all EURICA and DUTY children (the last participant of DUTY
43
44 432 study was recruited in April 2012). We will be able to describe serious short-term (30 days
45
46 433 and less than 1 year post index consultation) and medium-term (1-5 years) outcomes,
47
48 434 including hospital admission, renal imaging, renal scarring, VUR and renal failure outcomes
49
50 435 for all children in this group using PEDW data in Wales and NHS Digital hospital data in
51
52 436 England. We will be able to describe other outcomes, such as GP consultations, recurrent
53
54 437 UTI, dysfunctional voiding, antibiotic prescriptions, hypertension, chronic kidney disease,
55
56 438 using GP data on a sub-set. In Wales, Datastore will also be used to look at the urine culture
57
58 439 results and organism resistance profile for subsequent UTIs.
59
60

1
2
3 440 We will describe GP diagnosis from study data versus Read codes and acute management
4
5 441 from the routine data in GP records for this cohort for later comparisons and also to explore
6
7 442 the validity of using routinely collected data in these cases. We will also assess the validity of
8
9 443 using Read codes to diagnose UTI against microbiological culture results and agreement will
10
11 444 be measured using the Kappa statistic.
12

13
14 445

15
16
17 446 Research question 2: Comparison of short- and medium-term outcomes in children with
18
19 447 mcUTI: routine versus systematic sampling.
20

21
22 448 We will compare the outcomes in children with mcUTI identified through routine versus
23
24 449 systematic sampling. Children's characteristics, presentation factors, acute management
25
26 450 and microbiology results will be described for the groups using appropriate summary
27
28 451 statistics. We will compare urine sampling and UTI diagnosis in consultations between
29
30 452 routine and systematic sampling. In addition, we will describe blood pressure and creatinine
31
32 453 levels for each group if recorded and explore whether comparisons can be made.
33

34
35 454 Previously mentioned short- and medium-term outcomes will be described by the two groups
36
37 455 of routine vs. selective sampling. Predictors of outcome will be examined as before using a
38
39 456 multilevel multinomial regression model (no event, 30 days, event 30 days -1 year, event 1-5
40
41 457 years) and again where numbers allowed, variation in outcome will be accounted for at the
42
43 458 level of the general practice. Associations between covariates previously described and
44
45 459 outcome will firstly be examined. Unadjusted and adjusted relative risk ratios will be
46
47 460 estimated, together with 95% CIs. We will compare blood pressure and creatinine levels
48
49 461 (where available) across the groups; we expect this data to be limited so will be exploratory.
50

51
52
53 462 A detailed statistical analysis plan will be written prior to database lock. The reporting and
54
55 463 presentation of results will be in accordance with the [25–27] statements to ensure the
56
57 464 comprehensive reporting of our observational non-randomized evaluation of a public health
58
59 465 intervention. SPSS and Stata will be used for all analyses [28,29].
60

466 **Patient and Public Involvement**

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

475 **ETHICS AND DISSEMINATION**

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

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

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

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

504 **DECLARATIONS**

505 *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

506

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

511 *Consent for publication* - Not Applicable

512 *Availability of data and material* - Not Applicable

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

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

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

521 **Acknowledgements**

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

535

536 **References**

- 537 1 Craig JC, Williams GJ, Jones M, *et al*. The accuracy of clinical symptoms and signs
538 for the diagnosis of serious bacterial infection in young febrile children: prospective
539 cohort study of 15 781 febrile illnesses. *BMJ* 2010;**340**:c1594.
540 doi:10.1136/BMJ.C1594
- 541 2 Hoberman A, Chao HP, Keller DM, *et al*. Prevalence of urinary tract infection in febrile
542 infants. *J Pediatr* 1993;**123**:17–23.
- 543 3 Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and
544 economic costs. *Am J Med* 2002;**113 Suppl 1A**:5S–13S.
- 545 4 Hay AD, Birnie K, Busby J, *et al*. The Diagnosis of Urinary Tract infection in Young
546 children (DUTY): a diagnostic prospective observational study to derive and validate a
547 clinical algorithm for the diagnosis of urinary tract infection in children presenting to
548 primary care with an acute illness. *Health Technol Assess (Rockv)* 2016;**20**.
549 doi:10.3310/hta20510

- 1
2
3 550 5 O'Brien K, Edwards A, Hood K, *et al.* Prevalence of urinary tract infection in acutely
4 551 unwell children in general practice: a prospective study with systematic urine
5 552 sampling. *Br J Gen Pract* 2013;**63**:e156-64. doi:10.3399/bjgp13X663127
6
7
8 553 6 Urinary tract infection in under 16s: diagnosis and management | Guidance and
9 554 guidelines | NICE.
10
11
12 555 7 Butler CC, O'Brien K, Pickles T, *et al.* Childhood urinary tract infection in primary care:
13 556 a prospective observational study of prevalence, diagnosis, treatment, and recovery.
14 557 *Br J Gen Pract* 2015;**65**:e217-23. doi:10.3399/bjgp15X684361
15
16
17 558 8 Shaikh N, Ewing AL, Bhatnagar S, *et al.* Risk of Renal Scarring in Children With a
18 559 First Urinary Tract Infection: A Systematic Review. *Pediatrics* 2010;**126**:1084–91.
19 560 doi:10.1542/peds.2010-0685
20
21
22 561 9 Coulthard MG, Lambert HJ, Keir MJ. Occurrence of renal scars in children after their
23 562 first referral for urinary tract infection. *BMJ* 1997;**315**:918–9.
24
25
26 563 10 Jacobson SH, Eklöf O, Eriksson CG, *et al.* Development of hypertension and uraemia
27 564 after pyelonephritis in childhood: 27 year follow up. *BMJ* 1989;**299**:703–6.
28
29
30 565 11 Round J, Fitzgerald AC, Hulme C, *et al.* Urinary tract infections in children and the risk
31 566 of ESRF. *Acta Paediatr* 2012;**101**:278–82. doi:10.1111/j.1651-2227.2011.02542.x
32
33
34 567 12 Smellie JM, Prescod NP, Shaw PJ, *et al.* Childhood reflux and urinary infection: a
35 568 follow-up of 10-41 years in 226 adults. *Pediatr Nephrol* 1998;**12**:727–36.
36
37
38 569 13 Craig JC, Williams GJ. Denominators do matter: it's a myth--urinary tract infection
39 570 does not cause chronic kidney disease. *Pediatrics* 2011;**128**:984–5.
40 571 doi:10.1542/peds.2011-2631
41
42
43 572 14 Salo J, Ikäheimo R, Tapiainen T, *et al.* Childhood urinary tract infections as a cause of
44 573 chronic kidney disease. *Pediatrics* 2011;**128**:840–7. doi:10.1542/peds.2010-3520
45
46
47 574 15 Coulthard MG, Vernon SJ, Lambert HJ, *et al.* A nurse led education and direct access
48 575 service for the management of urinary tract infections in children: prospective
49 576 controlled trial. *BMJ* 2003;**327**:656. doi:10.1136/bmj.327.7416.656
50
51
52 577 16 Jadresic L, Cartwright K, Cowie N, *et al.* Investigation of urinary tract infection in
53 578 childhood. *BMJ* 1993;**307**:761–4. doi:10.1136/BMJ.307.6907.761
54
55
56 579 17 van der Voort J, Edwards A, Roberts R, *et al.* The struggle to diagnose UTI in children
57 580 under two in primary care. *Fam Pract* 1997;**14**:44–8.
58
59
60

- 1
2
3 581 18 Hollingworth W, Busby J, Butler CC, *et al.* The Diagnosis of Urinary Tract Infection in
4 582 Young Children (DUTY) Study Clinical Rule: Economic Evaluation. *Value Heal*
5 583 2017;**20**:556–66. doi:10.1016/j.jval.2017.01.003
6
7
8 584 19 Jones KH, Ford D V., Jones C, *et al.* A case study of the Secure Anonymous
9 585 Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for
10 586 health-related research and evaluation. *J Biomed Inform* 2014;**50**:196–204.
11 587 doi:10.1016/j.jbi.2014.01.003
12
13
14
15 588 20 Ford D V, Jones KH, Verplancke J-P, *et al.* The SAIL Databank: building a national
16 589 architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;**9**:157.
17 590 doi:10.1186/1472-6963-9-157
18
19
20
21 591 21 Lyons RA, Jones KH, John G, *et al.* The SAIL databank: linking multiple health and
22 592 social care datasets. *BMC Med Inform Decis Mak* 2009;**9**. doi:10.1186/1472-6947-9-3
23
24
25 593 22 Robling M, Bekkers M-J, Bell K, *et al.* Effectiveness of a nurse-led intensive home-
26 594 visitation programme for first-time teenage mothers (Building Blocks): a pragmatic
27 595 randomised controlled trial. *Lancet* 2016;**387**:146–55. doi:10.1016/S0140-
28 596 6736(15)00392-X
29
30
31
32 597 23 Public Health England. UK Standards for Microbiology Investigations. Investigation of
33 598 Urine. B41. 2018. [https://www.gov.uk/government/publications/smi-b-41-investigation-](https://www.gov.uk/government/publications/smi-b-41-investigation-of-urine)
34 599 [of-urine](https://www.gov.uk/government/publications/smi-b-41-investigation-of-urine)
35
36
37 600 24 Green SB. How Many Subjects Does It Take To Do A Regression Analysis.
38 601 *Multivariate Behav Res* 1991;**26**:499–510. doi:10.1207/s15327906mbr2603_7
39
40
41 602 25 Gilbert R, Lafferty R, Hagger-Johnson G, *et al.* GUILD: GUIDance for Information
42 603 about Linking Data sets†. *J Public Health (Bangkok)* 2017;**1**–8.
43 604 doi:10.1093/pubmed/fdx037
44
45
46 605 26 von Elm E, Altman DG, Egger M, *et al.* Strengthening the Reporting of Observational
47 606 Studies in Epidemiology (STROBE) statement: guidelines for reporting observational
48 607 studies. *BMJ* 2007;**335**:806–8. doi:10.1136/bmj.39335.541782.AD
49
50
51
52 608 27 Benchimol EI, Smeeth L, Guttman A, *et al.* The REporting of studies Conducted
53 609 using Observational Routinely-collected health Data (RECORD) Statement. *PLoS*
54 610 *Med* 2015;**12**:e1001885. doi:10.1371/journal.pmed.1001885
55
56
57 611 28 IBM. IBM SPSS Statistics for Windows: Version 22.0. 2013.
58
59
60 612 29 StataCorp. Stata Statistical Software: Release 15. 2017. 2017. doi:10.2307/2234838

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

614

For peer review only

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

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

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

618

619 (Title) Figure 2. Flow chart of study participants

620

621

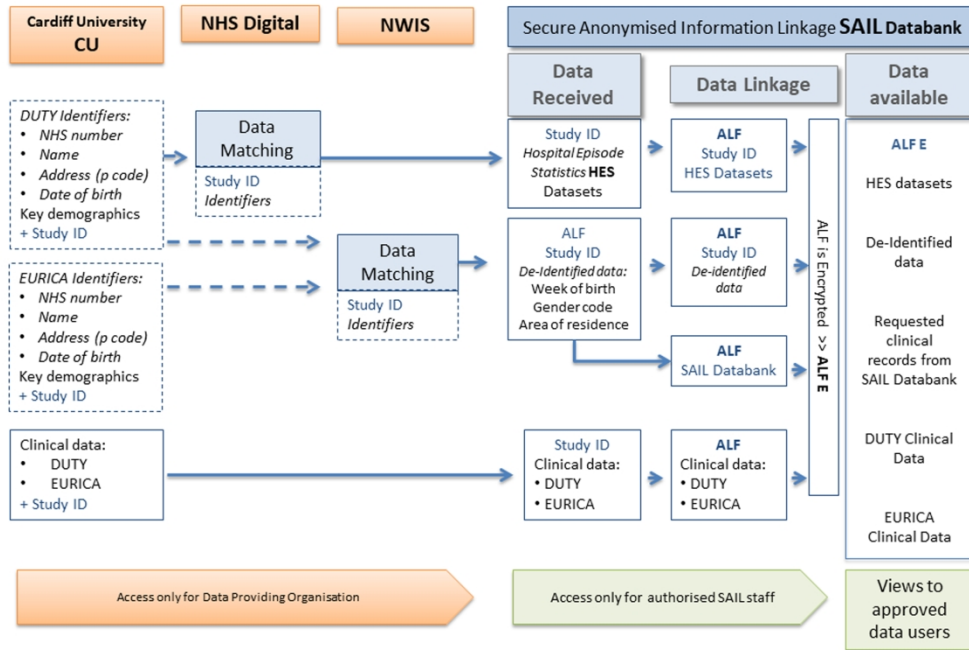
622

623

624

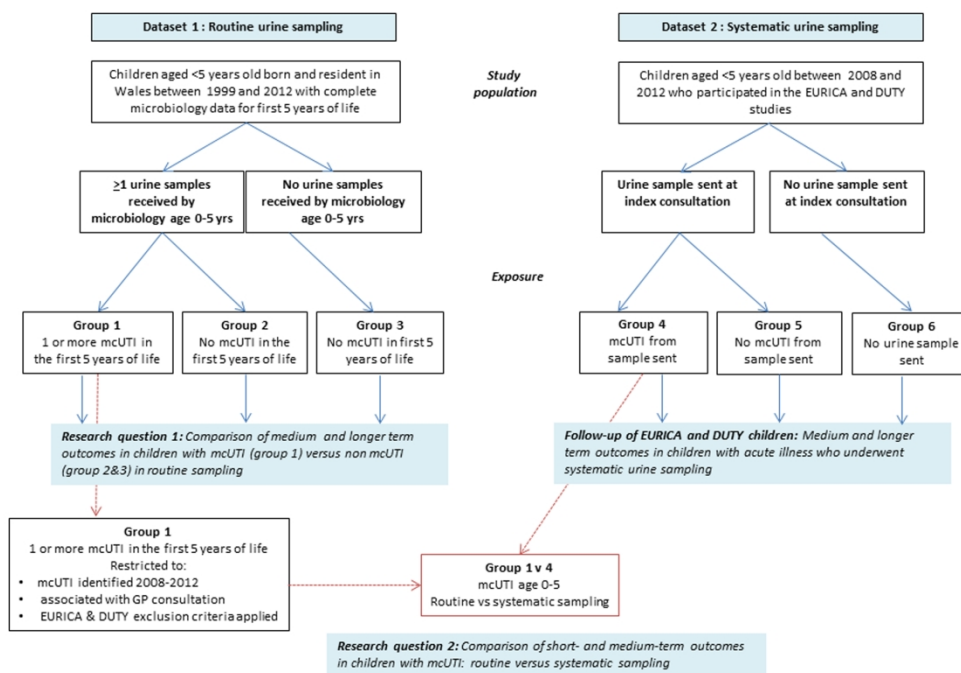
625

For peer review only



(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:	<i>BMJ Open</i>
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 Heginbotham, 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

SCHOLARONE™
Manuscripts

1
2
3 1 Long-term outcomes of urinary tract infection (UTI) in Childhood (LUCI): protocol for an
4 2 electronic record-linked cohort study
5
6 3
7

8 4 **Corresponding author:**

9 5 Fiona Lugg-Widger LuggFV@cardiff.ac.uk

10 6 Address: 4th Floor Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4YS. Centre for Trials
11 7 Research, College of Biomedical & Life Sciences, Cardiff University, Cardiff, UK.
12
13 8

14
15 9
16 10 Lianna Angel¹ AngelL@cardiff.ac.uk

17 11 Rebecca Cannings-John¹ CanningsRL@cardiff.ac.uk

18 12 Hywel Jones² JonesH75@cardiff.ac.uk

19 13 Mandy Lau¹ LauTM@cardiff.ac.uk

20 14 Christopher C Butler³ christopher.butler@phc.ox.ac.uk

21 15 Nick Francis⁴ FrancisNA@cardiff.ac.uk

22 16 Alastair D Hay⁵ alastair.hay@bristol.ac.uk

23 17 Margaret Heginbothom⁶ margaret.heginbothom@wales.nhs.uk

24 18 Kerenza Hood¹ HoodK1@cardiff.ac.uk

25 19 Shantini Paranjothy² ParanjothyS@cardiff.ac.uk

26 20 Judith Van der Voort⁷ judith.vandervoort@wales.nhs.uk

27 21 Kathryn Hughes⁴ HughesKA6@cardiff.ac.uk
28
29 22

30 23 **Author Affiliations**

- 31
32 24 1. Centre for Trials Research, College of Biomedical & Life Sciences, Cardiff University,
33 25 Cardiff, UK.
34 26 2. National Centre for Population Health and Wellbeing Research, Division of
35 27 Population Medicine, Cardiff University
36 28 3. Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe
37 29 Primary Care Building, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2
38 30 6GG
39
40 31 4. Wales Centre for Primary and Emergency Care Research (PRIME Centre Wales),
41 32 Division of Population Medicine, School of Medicine, Cardiff University
42 33 5. Centre of Academic Primary Care, NIHR School for Primary Care Research, Bristol
43 34 Medical School: Population Health Sciences, Canynge Hall, 39 Whatley Road, Bristol
44 35 BS8 2PS
45 36 6. Public Health Wales, 2 Capital Quarter, Tyndall Street, Cardiff CF10 4BZ
46 37 7. Noah's Arc Childrens Hospital for Wales, Cardiff, CF14 4XW
47
48
49
50 38
51
52 39
53
54 40

55 41 **Word Count:** 6810

56
57 42 **Keywords:** Children; Urinary Tract Infection; Medical record linkage; Renal scarring; Urine
58 43 sampling; Primary care.
59
60 44

1
2
3 45 **ABSTRACT**
4
5

6 46 **Introduction:** Current guidelines advise the prompt diagnosis and treatment of UTI in children
7
8 47 to improve both short and longer term outcomes. However, the risk of long-term complications
9
10 48 following childhood UTI is unclear.

11
12
13 49 UTI is relatively common but difficult to diagnose in children as symptoms are non-specific.
14
15 50 Diagnosis requires a urine sample, but sampling is difficult and infrequent, and it is not clear
16
17 51 if sampling should be given greater priority in primary care. The LUCI study will assess the
18
19 52 short, medium and longer-term outcomes of childhood UTI associated with routine and
20
21 53 systematic sampling practices.

22
23
24 54 **Methods and analysis:** Two datasets will be established: The first will consist of routinely
25
26 55 collected data (Hospital, GP, Microbiology) from children born and resident in Wales, linked
27
28 56 via the Secure Anonymised Information Linkage (SAIL) databank (an 'e-cohort'). Urine
29
30 57 sampling in this dataset reflects normal practice 'routine sampling'. Outcomes (including renal
31
32 58 scarring, hypertension, end-stage renal failure(ESRF), hospital admissions, GP consultations,
33
34 59 antibiotic prescriptions) for children with at least one UTI confirmed with microbiological culture
35
36 60 (mcUTI) or no mcUTI before the age of 5 will be compared.

37
38
39 61 The second will combine data from two prospective observational studies('DUTY' & 'EURICA')
40
41 62 employing systematic urine sampling for children presenting to primary care with acute,
42
43 63 undifferentiated illness, linked to routine data via SAIL (Wales) and NHS Digital (England).
44
45 64 Outcomes (as above, plus features of mcUTI) for children with a mcUTI in this dataset,
46
47 65 identified through systematic urine sampling, will be compared to those with a mcUTI identified
48
49 66 through routine urine sampling (dataset 1).
50
51

52
53 67 **Ethics and dissemination:** The study protocol has been approved by NHS Wales Research
54
55 68 Ethics Committee and the Health Research Authority's Confidentiality Advisory Group.
56
57 69 Methods of innovative study design and findings will be disseminated through peer-review
58
59
60

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

72 **Strengths and limitations of this study:**

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

85 **INTRODUCTION**

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

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

1
2
3 95 of renal scarring following first childhood UTI was 15%. [8] Most included studies were
4
5 96 conducted in secondary care and most required fever for inclusion in the study [8]. These are
6
7 97 likely to represent more serious UTIs than those presenting in primary care, and this rate of
8
9 98 renal scarring may not apply to all children presenting with UTI. The risk of long-term
10
11 99 complications (such as hypertension and renal failure) as a result of renal scarring following
12
13
14 100 childhood UTI is also not clear, with some researchers questioning this association. [6,13,14]
15
16 101 A review, commissioned by the National Institute for Health and Clinical Excellence (NICE),
17
18 102 concluded that 'there are no appropriate studies that accurately estimate the risks of long term
19
20 103 complications as a result of childhood UTI'. [6]

21
22
23 104 A urine sample is necessary to confirm the presence of UTI in childhood as the presenting
24
25 105 symptoms are non-specific and similar to those found in many common childhood illnesses
26
27 106 (e.g. respiratory infections, gastroenteritis, tonsillitis, ear infections) such as high temperature,
28
29 107 poor feeding, vomiting, lethargy and irritability. [6] Furthermore, microbiological confirmation
30
31 108 is important as some children with UTI require invasive investigations. [6] Children who
32
33 109 become seriously ill and who are assessed in emergency departments or admitted to hospital
34
35 110 will usually have their urine sampled. [1] However, in primary care, where most acute illness
36
37 111 is seen, urine is infrequently sampled. It is estimated that urine is sampled in fewer than 2%
38
39 112 of acute illness consultations with children under five years old in the UK. [4] Studies have
40
41 113 suggested that many cases of childhood UTI are missed in primary care. [15–17] Routine
42
43 114 practice (urine sampling based on clinician suspicion) is likely to miss more than two thirds of
44
45 115 UTIs in primary care. [5,7] The clinical implications of missed childhood UTIs are not known.
46
47 116 Increasing urine sampling in primary care is likely to increase the diagnosis of childhood
48
49 117 mcUTI, and is advocated by current guidelines, but it is not clear whether this is an appropriate
50
51 118 strategy, whether outcomes for children would improve or to what extent it should be
52
53
54 119 increased. [4–6,15,18]
55
56
57
58
59
60

1
2
3 120 Clarification is needed on two issues: First, to determine what the longer-term outcomes are
4
5 121 following childhood UTI (including those identified in primary care), and second, to determine
6
7 122 whether outcomes vary according to sampling practice.
8
9

10 123 In this study we will describe clinical outcomes for all children with one or more mcUTI aged
11
12 124 less than five years old, compared to those with no mcUTI, using NHS laboratory data from
13
14 125 across Wales. We will examine the risk factors for being diagnosed with renal scarring
15
16 126 following mcUTI.
17
18

19 127 We will also describe longer-term follow up of clinical outcomes for at least five years following
20
21 128 participation in two UK prospective cohort studies of acutely ill children with systematic urine
22
23 129 sampling in primary care, the DUTY and EURICA studies. [5,7] We will compare the outcomes
24
25 130 of those with mcUTI identified through these studies (systematic urine sampling) with those
26
27 131 identified through routine practice.
28
29

30 132 **METHODS AND DESIGN**

31 133 **Research objectives**

32
33
34
35
36 134 The LUCI Study will use data linkage of routinely collected datasets and data from two cohorts
37
38 135 of participants to answer two main research questions:
39
40

41 136 Research Question 1: Through routine sampling, do children who have experienced a mcUTI
42
43 137 aged <5 years old have worse outcomes (medium (0-5 years); longer-term (>5 years))
44
45 138 compared to children who have not experienced a mcUTI?
46
47

48 139 Research Question 2: Are short-term (<12 months) and medium-term (1-4 years) outcomes
49
50 140 different for children with childhood mcUTI identified through systematic sampling compared
51
52 141 to routine sampling (standard, clinician-led sampling)?
53
54

55 142 **Study Design**

56
57 143 This is a data linkage study comprising two overarching datasets of children. Dataset 1 will
58
59 144 comprise routinely collected health data from children born and resident in Wales. Children in
60

1
2
3 145 this dataset will have had urine sampled according to routine practice. Routine data will be
4
5 146 available on all children for seven years, and longer for some (i.e. children will be followed up
6
7 147 until the date of data abstraction). Dataset 1 will be used to answer Research Question 1.
8
9

10 148 Dataset 2 will be children who participated in the EURICA or DUTY studies. Children in this
11
12 149 dataset had their urine systematically sampled (all children presenting with an acute illness
13
14 150 were asked to provide a urine sample). DUTY and EURICA children will be followed-up by
15
16 151 linking records to routinely collected health data from England (using NHS Digital) and Wales
17
18 152 (using SAIL). Follow-up will be for a minimum of five years, and longer where data is available.
19
20 153 Dataset 2 will be used to describe longer term follow-up for DUTY and EURICA study children.
21
22 154 Those with mcUTI in Dataset 2 will be compared to those with mcUTI identified in Dataset 1
23
24 155 to answer Research Question 2.
25
26

27 156 The study formally started in October 2016 and will report to funder in June 2019. A summary
28
29 157 of the data sources is provided in Table 1.
30
31

32 158
33
34 159
35
36 160
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

161 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

For peer review only

163 **Data providers and datasets**

164 The EURICA and DUTY Studies

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

178

179 SAIL Databank

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

188

1
2
3 189 NHS Digital
4

5
6 190 NHS Digital is the trading name of the Health and Social Care Information Centre (HSCIC).
7

8 191 This study will access Hospital Episode Statistics (HES) data for participants of the DUTY and
9

10 192 EURICA study. All available Inpatient and Outpatient records belonging to each study
11

12 193 participant will be requested and approved by the Independent Group Advising on the Release
13

14 194 of Data (IGARD) Panel. The data requested includes diagnosis, procedures and length of
15

16 195 episode according to the 10th revision of the International Statistical Classification of Diseases
17

18 196 and Related Health Problems [ICD-10] codes.
19

20
21 197
22

23
24 198 Public Health Wales
25

26
27 199 Public Health Wales will provide a data extract of urine microbiology culture results from all
28

29 200 microbiology laboratories in Wales (Datastore) for use with this project. This will be transferred
30

31 201 to SAIL.
32

33 202
34

35 203 Individual Health Boards
36

37
38 204 Health boards in Wales will be approached to access anonymised radiology data for patients
39

40 205 in dataset 1. A one off data extract of patient-level attendance data for patients born between
41

42 206 01/01/1994 and 31/12/2012 that attended radiology between 1994 and 2016 will be
43

44 207 transferred to SAIL. Data extracted includes examination performed, attendance data and the
45

46 208 radiology report.
47

48
49 209
50

51
52 210 **Opportunity to opt-out (dataset two)**
53

54
55 211 Dataset one uses routinely collected data that is fully anonymised so we do not require
56

57 212 individual consent in order to access these data. Dataset two involves participants from the
58

59 213 DUTY and EURICA study and therefore requires section 251 (s251) support of the 2006 NHS
60

1
2
3 214 Act approval from the Health Research Authority's Confidentiality Advisory Group (HRA CAG)
4
5 215 to pass identifiable participant data legally held by Cardiff and Bristol University (as sponsors
6
7 216 of the studies) to the data providers. This is using an opt-out/dissent model instead of obtaining
8
9 217 further consent. In order to provide the opportunity for participants to dissent a letter has been
10
11 218 developed to be sent to all parents of participants. Participants have the opportunity to contact
12
13 219 the study team through an online web-form, by email, text or telephone (details provided on
14
15 220 website) and register their dissent. These participants who register their dissent will not be
16
17 221 included in any of the datasets sent to the information centres and therefore will not appear in
18
19 222 the datasets for analysis. A participant representative was consulted on the layout, wording
20
21 223 and level of information contained in the participant opt-out letter and on the study website. A
22
23 224 key consideration was to communicate the data transfer process. The final letter was approved
24
25 225 by both an NHS Research Ethics Committee and CAG committee as part of overall
26
27 226 governance approval for the study.
28
29
30

31
32
33

34 228 **Data Matching**

35
36
37 229 For included data sets, data will be sent to NHS Digital and SAIL (via NWIS their trusted third
38
39 230 party (TTP)) for matching using a combination of NHS number, name, address and date of
40
41 231 birth. Matching with NHS Digital data will be by exact matching on NHS Number, Date of Birth
42
43 232 and Postcode. This has been conducted in other studies and achieved a high match rate
44
45 233 (99.6%) [22]. NWIS match using NHS Number, Date of Birth, Sex, Postcode. Data matching
46
47 234 by NHS Digital will be separate to matching conducted by NWIS on behalf of SAIL. We would
48
49 235 expect only a small number of participants matched to both English and Welsh NHS records
50
51 236 however there is the possibility of this for those using services across the border to their current
52
53 237 address.
54
55

56 238

59 239 **The anonymised dataset**

1
2
3 240 For DUTY and EURICA participants NHS Digital will retain the unique study ID assigned by
4
5 241 the DUTY/EURICA study, with the HES records after the matching process before it is sent to
6
7 242 SAIL Databank. This ID will be retained for data transfer to ensure the data can be linked. The
8
9 243 same applies to data sent to NWIS. Requested data will then be linked for each participant via
10
11 244 SAIL using the unique study ID. The DUTY and EURICA clinical datasets will be transferred
12
13 245 to SAIL following a process of de-identification. The data flow is shown in Figure 1. The study
14
15 246 ID will be replaced by an anonymised linking field (ALF) to enable all datasets to be linked at
16
17 247 an individual level. The resulting dataset will contain clinical variables (from DUTY, EURICA,
18
19 248 SAIL and NHS Digital), de-identified demographic variables (e.g. week of birth and lower super
20
21 249 output area - LSOA) and the ALF. The key between study ID and ALF will be retained and
22
23 250 encrypted as a further safeguard. The ALF is further encrypted (ALF-E) so that datasets and
24
25 251 records cannot be linked across multiple projects a researcher has access to.
26
27
28
29
30

31 **Study Participants**

32
33
34 254 A flow chart of participants in the study is shown in Figure 2. Dataset 1 (routine sampling) will
35
36 255 be children born and resident in Wales for the first 5 years of life; who were less than 5 years
37
38 256 old between 1999 and 2012. The main analysis will be on children born between 01/01/2005
39
40 257 and 31/12/2009 to ensure that all of the first five years of life are covered by the dates when
41
42 258 Datastore is available.
43
44

45 259 Dataset 2 (systematic sampling from DUTY and EURICA) will be children who participated in
46
47 260 the EURICA or DUTY studies, who were not withdrawn from the study and when provided with
48
49 261 the opportunity to opt-out, remained in the study. For research question 2 when comparing
50
51 262 children with a mcUTI from Dataset 1 and 2, children in Group 1 will be selected to match the
52
53 263 DUTY and EURICA study eligibility criteria as closely as possible within the constraints of
54
55 264 using routine data. Datasets 1 and 2 will be limited to children with a mcUTI associated with a
56
57 265 GP consultation between March 2008 and April 2012. For both datasets, the child's mcUTI,
58
59 266 associated with a GP consultation (defined as within 14 days prior to the sample) will be
60

1
2
3 267 identified and defined as the index consultation. To limit the potential transfer of GP
4
5 268 systematic sampling behaviour, children from Group 1 with index consultations between 2008
6
7 269 and 2012 at practices which participated in the EURICA study will be flagged as will children
8
9 270 with index consultations between 2010 and 2012 at practices which participated in the DUTY
10
11 271 study. Children will only be included once in each study period (i.e. a child with a sample sent
12
13 272 within the EURICA study period could also appear in the DUTY study period). In addition, we
14
15 273 will apply the DUTY study exclusion criteria (where possible) to Dataset 1 to make them
16
17 274 directly comparable with Dataset 2; therefore children will be excluded from the routine
18
19 275 sampling cohort if any criterion met:
20
21
22

- 23 276
- 25 277 • Known neurogenic bladder (e.g. spina bifida) or previous bladder surgery
 - 27 278 • Prescribed antibiotics in the 7 days prior to presentation
 - 29 279 • Taking immunosuppressant medication
 - 31 280 • Using urinary catheters
- 33
34 281

35 282 Children with a prior history of UTI or vesicoureteric reflux (VUR) will not be excluded (and
36
37 283 were not excluded from the DUTY or EURICA studies) and we will explore the impact of these
38
39 284 risk factors on outcomes.
40
41

42 285

43 286
44 287

45 288 **Exposure**

47
48 289 UTI cases will be based on NHS laboratory results. For Dataset 1, this will be through
49
50 290 microbiological culture data downloaded from Datastore. These data represent samples (from
51
52 291 both community and hospital settings) which have been classified as positive or negative by
53
54 292 NHS laboratories according to their standard operating procedures. We do not know how urine
55
56 293 was sampled, and this is likely to vary between settings. In most cases, these are likely to be
57
58 294 clean catch samples, but may include urine collection pads or bags (particularly in community
59
60

1
2
3 295 samples) as recommended by NICE; or catheter or suprapubic aspiration (SPA) in hospital
4
5 296 samples.[6] NHS laboratories take into consideration the nature of the urine sample in their
6
7 297 reporting.[23] For Dataset 2, we will use the results of microbiological culture from NHS
8
9 298 laboratories collected during the DUTY and EURICA studies as some participants were from
10
11 299 England (Datastore is Wales only). For Dataset 2, the presence of significant bacteriuria (pure
12
13 300 or predominant growth of 100,000 cfu/ml of urine) will be used to define UTI.

14
15
16 301 For Dataset 1 we define the exposure period as <5 years and will be grouped as follows:
17
18 302 (Figure 1).

19
20
21 303 Group 1: children with at least one mcUTI before their 5th birthday or before outcome
22
23 304 of interest

24
25
26 305 Group 2: children with at least one urine sample but no mcUTI before their 5th birthday
27
28 306 or before outcome of interest

29
30
31 307 Group 3: children with no urine samples before their 5th birthday or before outcome of
32
33 308 interest

34
35
36 309 Exposure is a discrete time-varying covariate (0 (group 2 or 3 no UTI or no sample
37
38 310 respectively) until first exposure of group 1 (mcUTI) thereafter) and each patient's exposure
39
40 311 status will be taken at the point of each outcome; otherwise the exposure status of the child at
41
42 312 their 5th birthday will be taken.

43
44
45 313 For the main analyses, Groups 2 and 3 will be considered together as having no
46
47 314 microbiologically confirmed UTI. Dataset 2 will similarly be divided into three groups based on
48
49 315 their index consultation when recruited into the DUTY and EURICA studies:

50
51
52 316 Group 4: children with a mcUTI

53
54
55 317 Group 5: children who had a urine sample but no mcUTI

56
57
58 318 Group 6: children who had no urine sample
59
60

319

320

321 **Study variables**

322 Table 2 shows a breakdown of the baseline data and possible covariates available for children

323 and maternal characteristics from the data collection forms for EURICA and DUTY and WDS,

324 WECC, and for a subset with GP records. The study outcomes are summarised in Table 3.

325

326

327

328 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)		
<i>known to be associated</i> 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
<i>possibly associated</i> 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*
<i>Factors for follow-up of study participants & Research Question 2</i>		
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

329

330 Table 3. Study Outcomes

	Data source		
	PEDW (All Wales)	GP	Datastore (All Wales)
Primary outcome			
Renal scarring	✓		
<i>Sensitivity analyses</i>			
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)			✓

331

332 **Follow-up**

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

1
2
3 335 patient leaves the GP practice linked to SAIL or the last data collection from the general
4
5 336 practice. For the analysis of Research Question 1, using only children whose whole first five
6
7 337 years of life were covered by the dates that Datastore was available (excluding children whos
8
9 338 first five years of life fall outside of Datastore availability) and limiting DOBs to 1/1/2005 to
10
11 339 31/12/2009, will mean that shortest follow-up period will be 7 years. Follow-up will be longer
12
13 340 where data is available. For Research Question 2, we will examine outcomes at 30 days and
14
15 341 at 1 year (short term), 1-5 years (medium-term) and >5 years (long-term) after the index
16
17 342 consultation.

18
19
20 343
21 344 **Analysis**

22
23
24 345 **Sample size**

25
26
27 346 *Comparison of renal scarring in children routinely sampled for UTI with and without mcUTI:*

28
29 347 The sample size is based on the outcome of renal scarring of children with and without mcUTI
30
31 348 and taken as 15% [8]. Using an odds ratio of 1.5 (a conservative effect size of 5% difference
32
33 349 between groups), 2-sided alpha 0.05, power=90%, and given a ratio of 32:1 (children
34
35 350 diagnosed with mc UTI: children with no UTI diagnosis) in the dataset, we require 15,519
36
37 351 children (452 with mcUTI:15,067 without mcUTI) for analysis. Initial examination of the SAIL
38
39 352 dataset identified just under 13,000 children less than five years old with UTI between 1999
40
41 353 and 2012. As our sample size calculation requires 452 with UTI, we can be confident of
42
43 354 adequate power for this study. However, the true proportion with renal scarring is likely to be
44
45 355 less than that reported by Shaikh [8]. Assuming a proportion of 8% with renal scarring in the
46
47 356 mcUTI group, and an odds ratio of 1.5 (2.5% difference), 2-sided alpha=0.05, power=90%,
48
49 357 and a ratio of 32:1, 35,833 children (1,044 with UTI: 34,789 without UTI) are required for
50
51 358 analysis, which is still achievable.

52
53
54 359 *Comparison of systematically versus routinely sampled UTI:* This sample size is constrained
55
56 360 by the number of children with a systematically sampled microbiologically confirmed UTI by
57
58 361 NHS lab (N=374). Using the lower estimate of renal scarring of 8% of routinely sampled UTIs,
59
60

1
2
3 362 then a 5% difference (13% in the systematically sampled group) would give 89% power, with
4
5 363 a 2-sided alpha=0.05 and a ratio of 32:1 (routinely sampled UTIs 11,968: systematically
6
7 364 sampled UTIs 374).

8
9
10 365 Both analyses use multivariable regression. Using Green's [24] formulae, assuming medium
11
12 366 effect sizes (odds ratio=1.5), a 2-sided alpha=0.05, and power=80%, to test 20 predictors in
13
14 367 the multivariable regression model we require at least 159 children in total. This suggests that
15
16 368 we will be adequately powered for both analyses (given these assumptions) to examine
17
18 369 predictors of short and medium-term outcomes.

20
21 370

22
23
24 371 Statistical analysis

25
26
27 372 *Dataset 1: Routine sampling of UTI*

28
29
30 373 Research question 1: Comparison of medium (up to 5 years) and longer term (≥ 5 years)
31
32 374 outcomes in children with mcUTI versus those without mcUTI in routine sampling.

33
34 375 Baseline variables will be described using appropriate descriptive summaries (N (%), mean
35
36 376 (SD), median (interquartile range)) to summarise the population for the main analyses by
37
38 377 group 1, 2 and 3 and any marked imbalance between the groups will be identified. There will
39
40 378 be no formal testing of between-group differences for any variables at baseline. The main
41
42 379 comparative analyses will be carried out at a child level since outcomes relate to an individual's
43
44 380 exposure to one or more UTI and will test the null hypothesis that there are no differences in
45
46 381 outcomes in the first 5 years and between the age of 5 and 7 years. The primary analysis will
47
48 382 consist of a comparison of rate of diagnosis of renal scarring (no renal scarring, renal scarring
49
50 383 recorded <5 years, renal scarring recorded 5-7 years) in children with mcUTI versus children
51
52 384 with no mcUTI, using a multinomial regression model. Results will be reported as relative risk
53
54 385 ratios alongside 95% confidence intervals (CIs). A survival model will also be performed to
55
56 386 model time to first renal scarring diagnosis taking into account competing risks (such as deaths
57
58 387 and migration) and differences in time-at-risk and to allow us to look for this outcome using all

1
2
3 388 available follow-up for each child (at least 7 years). We will estimate hazard ratios with 95%
4
5 389 CIs for each exposure group.
6
7

8 390
9

10
11 391 We will construct a Direct Acyclic Graph (DAG) to inform our choice of variables to include in
12
13 392 the analysis. Confounding variables such as those listed in table 2 and also mcUTI that could
14
15 393 be considered to be on the causal pathway will be defined a priori. We will run multiple
16
17 394 mediation analyses using renal scarring as the dependent variable, mcUTI as the mediation
18
19 395 variable and confounders as the independent variables. First we will identify the independent
20
21 396 variables associated with renal scarring (using an univariable logistic (where scarring is rare)
22
23 397 or log-linear regression model (where scarring is common)) and identify the mediation
24
25 398 variables (mcUTI or not) that are associated with the significant independent variables. These
26
27 399 will all be included in the mediation model. For each of the significant independent variables,
28
29 400 two regression models will be performed with and without the mediation variable. We will
30
31 401 calculate the indirect effect (and the effect of the mediator) using the regression coefficients
32
33 402 from both regression models.
34
35

36
37 403 Unadjusted and adjusted relative risk ratios will be estimated, together with 95% CIs.
38

39
40 404 Several sensitivity analyses are proposed: The primary outcome will be expanded to include
41
42 405 any renal pathology codes due to uncertainty around whether the renal scarring codes are
43
44 406 sufficiently sensitive to pick up all cases of renal scarring. In addition we will explore using
45
46 407 Read codes from GP notes to pick up additional renal scarring diagnoses. Two effect modifiers
47
48 408 were identified as a basis for sub-group analyses for the primary outcome: gender of child and
49
50 409 presence of any renal/urological congenital anomalies. These pre-planned analyses will be
51
52 410 conducted by the inclusion of appropriate interaction terms in the models.
53

54
55 411 Secondary outcomes will be analysed using multinomial and time to event models (Table 3).

56
57 412 Poisson regression models will be used where the outcome is a count of event (e.g. hospital
58
59 413 admissions, GP consultations, antibiotics prescribed); results will be represented as incidence
60

1
2
3 414 rate ratios alongside 95% CIs. Additionally we will identify risk factors for renal scarring (as
4
5 415 above, plus age at first mcUTI, bacterial type & resistance) in children presenting with a
6
7 416 childhood mcUTI. In a sub sample of children linked to GP data, the 14 days prior to the urine
8
9 417 sample submission date will be examined to determine whether there was an associated GP
10
11 418 consultation. We will then estimate the likely rate of GP associated mcUTIs and the rate of
12
13 419 mcUTIs submitted from another setting (such as out of hours, outpatients or as a result of a
14
15 420 hospital admission). An identical analysis to the primary outcome will be taken to examine
16
17 421 whether the risk of renal scarring differs between those with a mcUTI or not, between different
18
19 422 settings and the interaction between the two. Where numbers allow, variation in outcome will
20
21 423 be accounted for at the level of the general practice in a multilevel model. We will also describe
22
23 424 the levels of urine sampling and incidence of mcUTI from General Practice in the routine data.
24
25

26 27 425 *Dataset 2: Systematic sampling of UTI*

28
29
30 426 Detailed study data for DUTY and EURICA participants is available including age, gender and
31
32 427 deprivation, presenting features, GP diagnosis and acute management. Recruited children are
33
34 428 already grouped into mcUTI or no mcUTI through the study microbiology data and outcomes
35
36 429 will be compared according to these groups. Five-year in-patient hospital outcomes will be
37
38 430 available for all EURICA and DUTY children (the last participant of DUTY study was recruited
39
40 431 in April 2012). We will be able to describe serious short-term (30 days and less than 1 year
41
42 432 post index consultation) and medium-term (1-5 years) outcomes, including hospital admission,
43
44 433 renal imaging, renal scarring, VUR and renal failure outcomes for all children in this group
45
46 434 using PEDW data in Wales and NHS Digital hospital data in England. We will be able to
47
48 435 describe other outcomes, such as GP consultations, recurrent UTI, dysfunctional voiding,
49
50 436 antibiotic prescriptions, hypertension, chronic kidney disease, using GP data on a sub-set. In
51
52 437 Wales, Datastore will also be used to look at the urine culture results and organism resistance
53
54 438 profile for subsequent UTIs.

55
56
57
58 439 We will describe GP diagnosis from study data versus Read codes and acute management
59
60 440 from the routine data in GP records for this cohort for later comparisons and also to explore

1
2
3 441 the validity of using routinely collected data in these cases. We will also assess the validity of
4
5 442 using Read codes to diagnose UTI against microbiological culture results and agreement will
6
7 443 be measured using the Kappa statistic.
8
9

10 444

11
12
13 445 Research question 2: Comparison of short- and medium-term outcomes in children with
14
15 446 mcUTI: routine versus systematic sampling.

16
17
18 447 We will compare the outcomes in children with mcUTI identified through routine versus
19
20 448 systematic sampling. Children's characteristics, presentation factors, acute management and
21
22 449 microbiology results will be described for the groups using appropriate summary statistics. We
23
24 450 will compare urine sampling and UTI diagnosis in consultations between routine and
25
26 451 systematic sampling. In addition, we will describe blood pressure and creatinine levels for
27
28 452 each group if recorded and explore whether comparisons can be made.

29
30
31 453 Previously mentioned short- and medium-term outcomes will be described by the two groups
32
33 454 of routine vs. selective sampling. Predictors of outcome will be examined as before using a
34
35 455 multilevel multinomial regression model (no event, 30 days, event 30 days -1 year, event 1-5
36
37 456 years) and again where numbers allowed, variation in outcome will be accounted for at the
38
39 457 level of the general practice. Associations between covariates previously described and
40
41 458 outcome will firstly be examined. Unadjusted and adjusted relative risk ratios will be estimated,
42
43 459 together with 95% CIs. We will compare blood pressure and creatinine levels (where available)
44
45 460 across the groups; we expect this data to be limited so will be exploratory.

46
47
48 461 A detailed statistical analysis plan will be written prior to database lock. The reporting and
49
50 462 presentation of results will be in accordance with the [25–27] statements to ensure the
51
52 463 comprehensive reporting of our observational non-randomized evaluation of a public health
53
54 464 intervention. SPSS and Stata will be used for all analyses [28,29].

55
56
57
58 465 **Patient and Public Involvement**
59
60

1
2
3 466 We have a parent representative (Sarah Jones) who has contributed to all stages of this study.
4
5 467 She helped to organise a parent group to discuss information provided to DUTY and EURICA
6
7 468 study participants explaining the study and opt-out mechanism. She also provided input on
8
9 469 the study website and on the procedure in place to manage contacts made by the participants.
10
11 470 During the drafting of the statistical analysis plan we discussed the planned analyses with her,
12
13 471 and she identified which of the analyses that she felt would be of most interest to parents of
14
15 472 children with suspected UTI. Results will be disseminated via the study website and other
16
17 473 channels with the input from our parent representative.
18
19
20

21 474 **ETHICS AND DISSEMINATION**

22
23 475 The governance surrounding dataset one differs from dataset two. Dataset one is an
24
25 476 anonymised dataset made available from SAIL databank with only approval required from the
26
27 477 IGRP whereas dataset two involves the transfer of identifiable data to data providers which
28
29 478 requires ethical approval, section 251 (s251) support, IGRP and IGARD approval. In order to
30
31 479 obtain an unbiased sample from the DUTY and EURICA cohort an opt-out consent model is
32
33 480 being used, this was supported by both the ethics panel and the CAG panel as justification for
34
35 481 this model of consent.
36
37
38

39 482 The LUCI study will report the risk of renal scarring for children with and without childhood
40
41 483 mcUTI across the whole of Wales. The linking of routinely collected datasets will give us a
42
43 484 large cohort including demographic, hospital in-patient and out-patient, GP and microbiology
44
45 485 data, allowing us to define mcUTI cases and describe outcomes for all children from both
46
47 486 primary and secondary care. Clarifying the link between UTI, renal scarring and long-term
48
49 487 complications will inform the management of acutely ill children in primary care, where the
50
51 488 need for urine sampling is unclear. Determining the clinical implications of 'missed' cases of
52
53 489 UTI through our comparison of children with mcUTI identified through routine and systematic
54
55 490 urine sampling will also help to determine the most appropriate urine sampling strategy. This
56
57 491 study maximises the benefits of the previously funded DUTY and EURICA cohorts,
58
59 492 representing over 8000 acutely ill children recruited from UK primary care. Significant
60

1
2
3 493 resources were invested by funders, patients and staff to develop these cohorts. Routine data
4
5 494 linkage will allow us to determine longer-term outcomes for these children and to determine
6
7 495 risks of adverse outcomes. It will also pave the way for even longer-term follow-up of cohorts
8
9 496 of children with UTI (diagnosed both systematically and routinely) which has been identified
10
11 497 as a high research priority by NICE.

12
13
14 498 A lay summary of the results and links to publications will be made available on the University
15
16 499 project website. The academic outputs for this study include (i) this protocol paper, (ii) main
17
18 500 results from research question one and (ii) main results from research question two. The
19
20 501 findings from this study will be of interest to clinicians and policy makers and may influence
21
22 502 the management of acutely ill children and childhood UTI.

503 **DECLARATIONS**

504 *List of abbreviations*

505 ALF: Anonymised linking field

506 ALF-E: Anonymised linking field encryption

507 CIs: Confidence intervals

508 DAG: Directed acyclic graph

509 DOB: Date of Birth

510 ESRF: End-stage renal failure

511 HES: Hospital Episode Statistics

512 HRA CAG: Health Research Authority's Confidentiality Advisory Group

513 HSCIC: Health and Social Care Information Centre

514 IGARD: Independent Group Advising on the Release of Data

515 IGRP: Information Governance Review Panel

516 LSOA: Lower super output area

517 mcUTI: Microbiological culture urinary tract infection

518 NICE: National Institute for Health and Clinical Excellence

519 NIHR HTA: National Institute of Health Research Health Technology Assessment

520 NISCHR: National Institute for Social Care and Health Research

521 NWIS: NHS Wales Informatics Service

522 PEDW: Patient Episode Database for Wales

1
2
3 523 PRIME: Primary and Emergency Care Research

4
5 524 SAIL: Secure Anonymised Information Linkage

6
7 525 TTP: Trusted third party

8
9 526 UTI: Urinary tract infection

10
11 527 VUR: Vesicoureteric reflux

12
13 528 WDS: Welsh Demographic Service

14
15 529 WECC: Welsh Electronic Cohort of Children

16
17 530

18 531 *Ethics approval and consent to participate* - Ethics approval of the study has been given by

19
20 532 the Research Ethics Committee for Wales (16/WA/0166) and the transfer and use of

21
22 533 identifiable data has been approved by the Health Research Authority [HRA] Confidentiality

23
24 534 Advisory Group [CAG] (16/CAG/0114).

25
26 535 *Consent for publication* - Not Applicable

27
28
29 536 *Availability of data and material* - Not Applicable

30
31
32 537 *Competing Interests* - The authors declare that they have no competing interests

33
34
35 538 *Funding* - This project has been funded by the Welsh Government through Health and Care

36
37 539 Research Wales [Project number 1068].

38
39
40 540 *Authors' contributions*- KHu is the chief investigator of the study. All authors have contributed

41
42 541 to and are responsible for the final design of the study. FLW is responsible for study

43
44 542 management. RCJ and ML are responsible for statistical planning and analysis. LA and HJ

45
46 543 are responsible for the data management. All authors have read and approved the final

47
48 544 manuscript [KHu, FLW, RCJ, ML, LA, HJ, CB, NF, AH, MH, KH, SP & JV] .

49
50
51 545 ***Acknowledgements***

52
53
54 546 We would like to acknowledge the support and input from Sarah Jones, our parent

55
56 547 representative for the study. We are also grateful to the DUTY and EURICA participants for

57
58 548 their agreement for continued use of their data for this study. The Centre for Trials Research

59
60

1
2
3 549 receives funding from Health and Care Research Wales and Cancer Research UK. Wales
4
5 550 Centre for Primary and Emergency Care Research (PRIME Centre Wales) receives funding
6
7 551 from Health and Care Research Wales. Authors are supported by The Farr Institute CIPHER,
8
9 552 funded by Arthritis Research UK, the British Heart Foundation, Cancer Research UK, the
10
11 553 Economic and Social Research Council, the Engineering and Physical Sciences Research
12
13 554 Council, the Medical Research Council, the National Institute of Health Research, the National
14
15 555 Institute for Social Care and Health Research (Welsh Assembly Government), the Chief
16
17 556 Scientist Office (Scottish Government Health Directorates), and the Wellcome Trust, (MRC
18
19 557 Grant No: MR/K006525/1) and the National Centre for Population Health & Wellbeing
20
21 558 Research (NCPHWR).
22
23
24
25 559

26 560 **References**

- 27
28
29 561 1 Craig JC, Williams GJ, Jones M, *et al*. The accuracy of clinical symptoms and signs
30 562 for the diagnosis of serious bacterial infection in young febrile children: prospective
31 563 cohort study of 15 781 febrile illnesses. *BMJ* 2010;**340**:c1594.
32 564 doi:10.1136/BMJ.C1594
33
34
35
36 565 2 Hoberman A, Chao HP, Keller DM, *et al*. Prevalence of urinary tract infection in febrile
37 566 infants. *J Pediatr* 1993;**123**:17–23.
38
39
40 567 3 Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and
41 568 economic costs. *Am J Med* 2002;**113 Suppl 1A**:5S–13S.
42
43
44 569 4 Hay AD, Birnie K, Busby J, *et al*. The Diagnosis of Urinary Tract infection in Young
45 570 children (DUTY): a diagnostic prospective observational study to derive and validate a
46 571 clinical algorithm for the diagnosis of urinary tract infection in children presenting to
47 572 primary care with an acute illness. *Health Technol Assess (Rockv)* 2016;**20**.
48 573 doi:10.3310/hta20510
49
50
51
52 574 5 O'Brien K, Edwards A, Hood K, *et al*. Prevalence of urinary tract infection in acutely
53 575 unwell children in general practice: a prospective study with systematic urine
54 576 sampling. *Br J Gen Pract* 2013;**63**:e156-64. doi:10.3399/bjgp13X663127
55
56
57 577 6 Urinary tract infection in under 16s: diagnosis and management | Guidance and
58 578 guidelines | NICE.
59
60

- 1
2
3 579 7 Butler CC, O'Brien K, Pickles T, *et al.* Childhood urinary tract infection in primary care:
4 a prospective observational study of prevalence, diagnosis, treatment, and recovery.
5 580
6 581 *Br J Gen Pract* 2015;**65**:e217-23. doi:10.3399/bjgp15X684361
7
- 8 582 8 Shaikh N, Ewing AL, Bhatnagar S, *et al.* Risk of Renal Scarring in Children With a
9 583 First Urinary Tract Infection: A Systematic Review. *Pediatrics* 2010;**126**:1084–91.
10 584 doi:10.1542/peds.2010-0685
11
- 12 585 9 Coulthard MG, Lambert HJ, Keir MJ. Occurrence of renal scars in children after their
13 586 first referral for urinary tract infection. *BMJ* 1997;**315**:918–9.
14
- 15 587 10 Jacobson SH, Eklöf O, Eriksson CG, *et al.* Development of hypertension and uraemia
16 588 after pyelonephritis in childhood: 27 year follow up. *BMJ* 1989;**299**:703–6.
17
- 18 589 11 Round J, Fitzgerald AC, Hulme C, *et al.* Urinary tract infections in children and the risk
19 590 of ESRF. *Acta Paediatr* 2012;**101**:278–82. doi:10.1111/j.1651-2227.2011.02542.x
20 591
- 21 591 12 Smellie JM, Prescod NP, Shaw PJ, *et al.* Childhood reflux and urinary infection: a
22 592 follow-up of 10-41 years in 226 adults. *Pediatr Nephrol* 1998;**12**:727–36.
23
- 24 593 13 Craig JC, Williams GJ. Denominators do matter: it's a myth--urinary tract infection
25 594 does not cause chronic kidney disease. *Pediatrics* 2011;**128**:984–5.
26 595 doi:10.1542/peds.2011-2631
27
- 28 596 14 Salo J, Ikäheimo R, Tapiainen T, *et al.* Childhood urinary tract infections as a cause of
29 597 chronic kidney disease. *Pediatrics* 2011;**128**:840–7. doi:10.1542/peds.2010-3520
30 598
- 31 598 15 Coulthard MG, Vernon SJ, Lambert HJ, *et al.* A nurse led education and direct access
32 599 service for the management of urinary tract infections in children: prospective
33 600 controlled trial. *BMJ* 2003;**327**:656. doi:10.1136/bmj.327.7416.656
34 601
- 35 601 16 Jadresic L, Cartwright K, Cowie N, *et al.* Investigation of urinary tract infection in
36 602 childhood. *BMJ* 1993;**307**:761–4. doi:10.1136/BMJ.307.6907.761
37 603
- 38 603 17 van der Voort J, Edwards A, Roberts R, *et al.* The struggle to diagnose UTI in children
39 604 under two in primary care. *Fam Pract* 1997;**14**:44–8.
40 605
- 41 605 18 Hollingworth W, Busby J, Butler CC, *et al.* The Diagnosis of Urinary Tract Infection in
42 606 Young Children (DUTY) Study Clinical Rule: Economic Evaluation. *Value Heal*
43 607 2017;**20**:556–66. doi:10.1016/j.jval.2017.01.003
44 608
- 45 608 19 Jones KH, Ford D V., Jones C, *et al.* A case study of the Secure Anonymous
46 609 Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for
47 610 health-related research and evaluation. *J Biomed Inform* 2014;**50**:196–204.

- 1
2
3 611 doi:10.1016/j.jbi.2014.01.003
4
5 612 20 Ford D V, Jones KH, Verplancke J-P, *et al.* The SAIL Databank: building a national
6 architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;**9**:157.
7 613 doi:10.1186/1472-6963-9-157
8 614
9
10 615 21 Lyons RA, Jones KH, John G, *et al.* The SAIL databank: linking multiple health and
11 social care datasets. *BMC Med Inform Decis Mak* 2009;**9**. doi:10.1186/1472-6947-9-3
12 616
13
14 617 22 Robling M, Bekkers M-J, Bell K, *et al.* Effectiveness of a nurse-led intensive home-
15 visitation programme for first-time teenage mothers (Building Blocks): a pragmatic
16 618 randomised controlled trial. *Lancet* 2016;**387**:146–55. doi:10.1016/S0140-
17 619 6736(15)00392-X
18 620
19
20 621 23 Public Health England. UK Standards for Microbiology Investigations. Investigation of
21 Urine. B41. 2018. [https://www.gov.uk/government/publications/smi-b-41-investigation-](https://www.gov.uk/government/publications/smi-b-41-investigation-of-urine)
22 622 of-urine
23 623
24
25 624 24 Green SB. How Many Subjects Does It Take To Do A Regression Analysis.
26 *Multivariate Behav Res* 1991;**26**:499–510. doi:10.1207/s15327906mbr2603_7
27 625
28
29 626 25 Gilbert R, Lafferty R, Hagger-Johnson G, *et al.* GUILD: GUIDance for Information
30 about Linking Data sets†. *J Public Health (Bangkok)* 2017;**;**1–8.
31 627 doi:10.1093/pubmed/idx037
32 628
33
34 629 26 von Elm E, Altman DG, Egger M, *et al.* Strengthening the Reporting of Observational
35 Studies in Epidemiology (STROBE) statement: guidelines for reporting observational
36 630 studies. *BMJ* 2007;**335**:806–8. doi:10.1136/bmj.39335.541782.AD
37 631
38
39 632 27 Benchimol EI, Smeeth L, Guttman A, *et al.* The REporting of studies Conducted
40 using Observational Routinely-collected health Data (RECORD) Statement. *PLoS*
41 633 *Med* 2015;**12**:e1001885. doi:10.1371/journal.pmed.1001885
42 634
43
44 635 28 IBM. IBM SPSS Statistics for Windows: Version 22.0. 2013.
45 636
46
47 636 29 StataCorp. Stata Statistical Software: Release 15. 2017. 2017. doi:10.2307/2234838
48
49
50
51
52
53
54 638
55
56
57
58
59
60

1
2
3 639 (Title) Figure 1. The data flow for dataset 2.
4
5

6 640 (Legend) ALF- Anonymised Linking Field; NWIS – NHS Wales Informatics Service; HES –
7
8 641 Hospital Episode Statistics; SAIL – Secure Anonymised Information Linkage
9

10
11 642
12

13
14 643 (Title) Figure 2. Flow chart of study participants
15

16 644
17

18 645
19

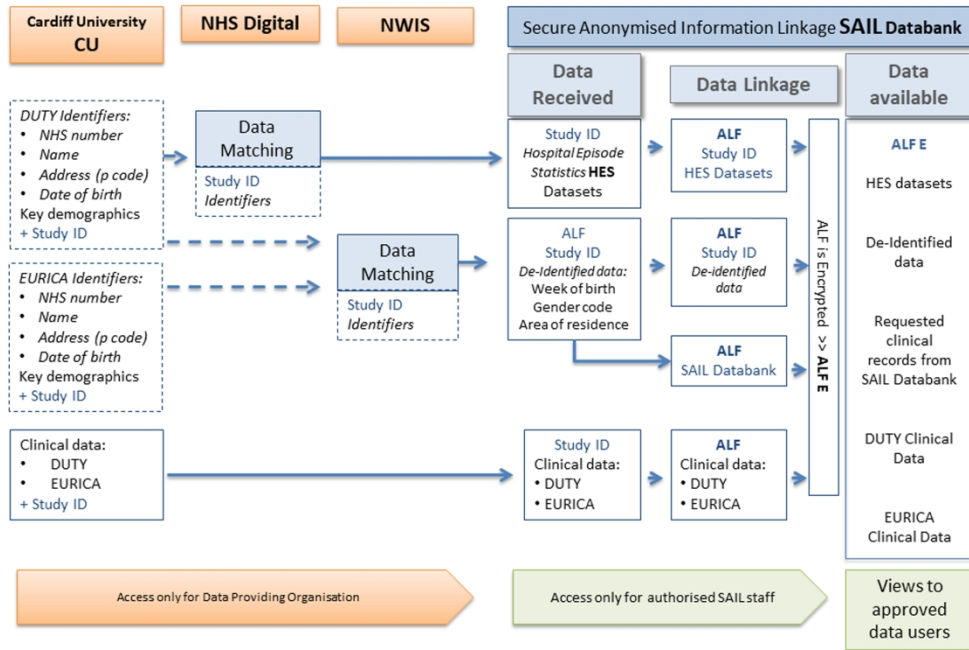
20 646
21

22 647
23

24 648
25

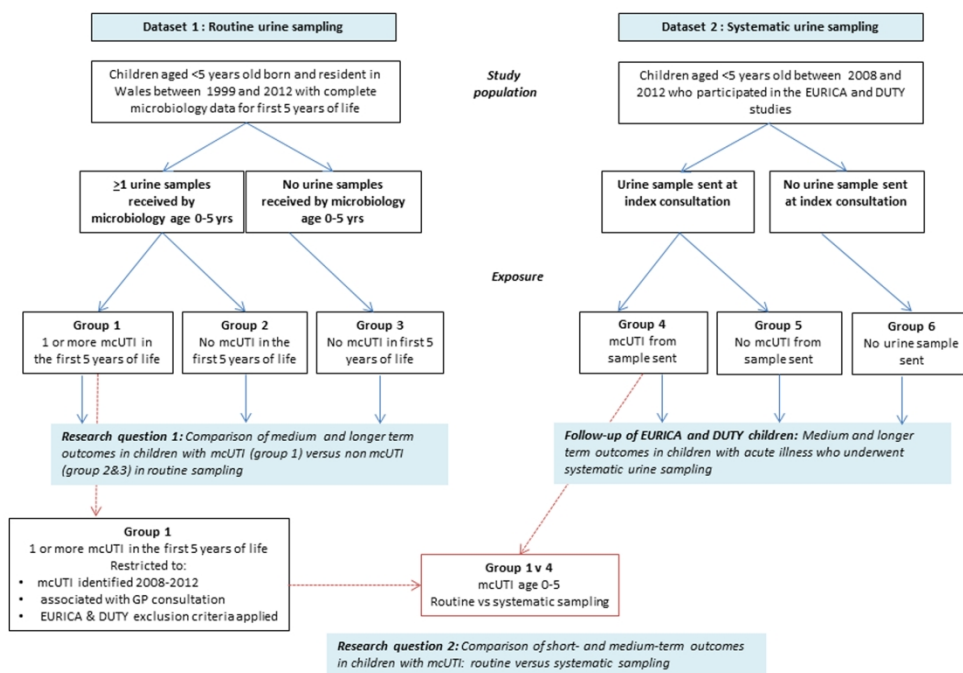
26 649
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only



(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)