1 **Manuscript Cover Sheet**

2

3

4

5 **Manuscript Title:** Epidemiology of distal renal tubular acidosis, a study using linked UK primary care and

6 hospital data

7

8

9 **Authors:** F. Bianic, F. Guelfucci, L. Robin, C. Martre, D. Game, D. Bockenhauer

10

11 **Version:** Draft 4

12

13 **Article Type:** Research

14

15

16

17 **Title:** Epidemiology of distal renal tubular acidosis, a study using linked UK primary care and hospital

18 data

19

20 **Authors:** Florence Bianic,[1] Florent Guelfucci,[1] Ludovic Robin,[2] Catherine Martre,[2] David Game,[3] Detlef

21 Bockenhauer[4]

22

23 **Affiliations:**

24 1. Syneos Health Consulting, Paris, France

25 2. Advicenne, Paris, France

26 3. Guy's and St Thomas' Hospital, London, England

27 4. Great Ormond Street Hospital, London, England

28

29 **Corresponding author:**

30 Florent Guelfucci

31 Email: florent.guelfucci@syneoshealth.com

32

33

Abstract

Word count: 348 (maximum: 350)

Background: Distal renal tubular acidosis (dRTA), or renal tubular acidosis (RTA) type 1, a rare inherited or acquired disease, is a disorder of the distal tubule. It is caused by impaired urinary acid secretion by the α-intercalated cells of the kidney with consequent systemic acidosis. Due to associated conditions and non-specific symptoms it may go undetected. This analysis aims to estimate the prevalence of dRTA in a large UK electronic medical record database and extrapolate to European Union Five (EU5) populations.

Methods: A retrospective analysis was carried out using the Clinical Practice Research Datalink (CPRD) GOLD UK database and linked Hospital Episode Statistics (HES) data to identify diagnosed and potentially undiagnosed or miscoded patients. A preliminary extraction of patients with at least one diagnosis code for dRTA, RTA, specific autoimmune diseases or renal disorders recorded between January 1987 and November 2017 were obtained from CPRD. Patients with a coded diagnosis of dRTA/RTA were analyzed on the following aspects: demographics, treatment and comorbidities. An algorithm was developed to detect potentially undiagnosed or uncoded dRTA, based on a sequence of inclusion criteria that included the presence of conditions known to be associated with inherited and acquired dRTA and prescriptions for alkali supplementation (suspected cases). Prevalence rates for 2017 were calculated and applied to EU5 populations.

Results: Two hundred and sixteen patients with a recorded diagnosis of RTA or dRTA were identified from the database, of whom 98 had a linkage to hospital data. A total of 447 patients were identified as having suspected dRTA through the algorithm. The dRTA prevalence was estimated to be between 0.46 (recorded diagnosed cases, of which 22.1% were considered primary [inherited]) and 1.60 if we include the suspected cases (of which 7.6% primary) per 10,000 people in 2017. Prescription and clinical records of diagnosed patients revealed a wide range of comorbidities (including renal conditions, anemia, hearing problems) and a need for pharmacological treatment to manage associated symptoms after diagnosis of dRTA.

Conclusions: The study provides new estimates of dRTA prevalence in Europe and suggests that patients may often be unreported or miscoded, potentially confounding appropriate disease management.

Keywords (3–10): dRTA, prevalence, epidemiology, diagnosed, misdiagnosed, miscoded, undiagnosed.

62 **Word count: 4,011**

63 Background

64 Renal tubular acidosis (RTA) is characterized by the buildup of acid in the body due to impaired urinary

65 acidification [1]. Distal RTA (dRTA) is mainly caused by defective H^+-ATPase or anion exchanger 1 (AE1)

66 transporters in the α-intercalated cells of the collecting duct. The typical clinical symptoms include

67 hyperchloremic metabolic acidosis, hypokalemia, hypocitraturia and hypercalciuria with consequent

68 nephrocalcinosis and/or nephrolithiasis [2, 3].

69

70 Buffering of the excess acid leads to decreased plasma HCO_3. In addition, acid is buffered by the bone with

71 release of bicarbonate and phosphate complexed with calcium stored in bone. The excess calcium released

72 from bone mineral resorption leads to high calcium excretion (hypercalciuria); Lastly, there is enhanced

73 reabsorption of filtered citrate in the proximal tubules with consequent decreased levels of citrate in the

74 urine (hypocitraturia).

75

76 Calcium bicarbonate and calcium phosphate efflux from the bone results in bone softening manifestations,

77 such as rickets or osteomalacia in children or juveniles/adults, respectively, osteoporosis and fractures [2,

78 4, 5]. The hypercalciuria, together with the hypocitraturia promotes abnormal renal calcium deposition such

79 as nephrocalcinosis and/or nephrolithiasis both of which may result in progressive chronic kidney disease

80 [6].

81

82 Further symptoms of dRTA include weakness, fatigue and low potassium levels in the patient's blood which

83 can lead to acute metabolic emergencies, cardiac arrhythmias, periodic paralysis, acute respiratory failure

84 and even sudden death.

85

86 dRTA can be inherited or acquired; therefore, onset can occur at any age depending on the underlying

87 cause [3]. However, inherited dRTA is less common than acquired dRTA [7]. Acquired dRTA is associated

88 with autoimmune diseases, such as Sjögren's syndrome or systemic lupus erythematosus (SLE), sickle cell

89 disease, chronic obstructive uropathy, or post-renal transplant [8]. Genes implicated in inherited dRTA

90   include *SLC4A1* [9] *ATP6V0A4* [7], *ATP6V1B1* [2], *WDR72* [3] and *FOXI1* [9]. Mutations in SLC4A1 are

91   usually dominant, but can be recessive, all other genes are associated with autosomal recessive

92   inheritance. In patients with autosomal recessive dRTA, disease onset typically occurs during infancy and

93   may be accompanied by sensorineural hearing loss, whereas in patients with autosomal dominant dRTA,

94   the diagnosis may be delayed until adolescence or young adulthood [3].

95

96   Treatment of dRTA aims to correct the biochemical abnormality of the disease. As of today, no treatment

97   has been approved for dRTA. Alkalizing treatment with citrate and/or bicarbonate complexed with

98   potassium and/or sodium in various non-approved formulations is used for patients with dRTA to buffer the

99   excess acid in the body [10, 11].

100

101   As dRTA is a rare disease, clinicians may be unfamiliar with the diagnosis. Consequently, there is a risk

102   that dRTA may be under-reported and miscoded, not least due to lack of specificity in the existing coding

103   systems (no ICD-10 code). The prevalence and incidence of dRTA are difficult to evaluate and published

104   studies typically concern patient-specific case studies rather than exploring the epidemiology of the disease

105   [11, 12].

106

107   In the UK, for rare non-emergency health issues, such as dRTA, patients seek follow-up consultation from

108   general practitioners (GPs) in primary healthcare. The Clinical Practice Research Datalink (CPRD), a UK-

109   based real-world research service, provides access to a large primary care database, the CPRD GOLD,

110   which contains longitudinal anonymized patient data (79 million person-years of follow-up), routinely

111   collected since 1987 across a network of approximately 674 GP practices in the UK (Herrett et al. 2015).

112   In this database, patient conditions are coded by general practice staff with an appropriate level of

113   granularity, providing a valuable source of patient-level data for epidemiological studies [13].

114

115   Understanding the prevalence of dRTA and treatments prescribed will provide a better insight into of the

116   unmet need and the challenges that patients and physicians face in terms of diagnosis and optimal

117   treatment. This study aims to estimate the 2017 prevalence of secondary (or acquired) and primary (or

118   inherited) dRTA in a primary care database CPRD GOLD in the UK and, by extrapolation, in the French,

119   German, Italian and Spanish general populations. Moreover, we describe the demographic and clinical

120   characteristics of these patients and identify the most frequently prescribed treatment.

121   In the study from Lopez-Garcia et al. [14], only 6 out of 336 patients with primary dRTA (1.6%) presented

122   in adulthood, suggesting that most inherited dRTA are detected before age 20 years. Conversely,

123   secondary dRTA (type 1) in the context of autoimmune disease presents almost exclusively in adults (Nat

124   Rev Nephrol. 2016 Feb;12(2):82-93. doi: 10.1038/nrneph.2015.174).

125   Consequently, for the purpose of this study, we defined dRTA with onset < 20 year of life as presumed

126   primary and $\geq$ 20 years as presumed secondary dRTA.

127

128   **Results**

129   **Diagnosed patients from the database**

130   <u>Demographics</u>

131   A total of 212 patients with coding for RTA and four patients with coding for dRTA were identified in the

132   CPRD GOLD database. Importantly, a careful review of their medical records showed that none of these

133   patients had features suggestive of renal Fanconi Syndrome/proximal RTA (type II). This is consistent with

134   dRTA being the most common form of RTA (90%) [11]; therefore, this study assumed that all patients coded

135   with RTA had dRTA.

136

137   Of the patients with a coded diagnosis of RTA/dRTA, a total of 98 patients were eligible for linkage to the

138   Hospital Episode Statistics (HES) data. Mean registration time was 22.8 years and 16.8 years in patients

139   with RTA and dRTA, respectively. The first patient with dRTA was diagnosed in 1958, and half of the

140   patients with dRTA were diagnosed after 2006. On average, 10 patients were newly diagnosed each year

141   from 2006 to 2017.

142

143   Mean age at diagnosis (standard deviation; SD) was 46 years (25.9 years), where the date of diagnosis

144   was defined as either the date of the first visit to a nephrologist or the date of the first record of RTA/dRTA

145   by the GP, whatever came first. This was decided based on the assumption that the dRTA diagnosis would

146   be carried out by a specialist rather than a GP and that GPs would then include that diagnosis in the

147   database. Based on our suggested age cut-off of 20 years (see methods), 18.0% of patients with a recorded

148   diagnosis in the CPRD GOLD database over the study period (January 1987 - November 2017) had primary

149   dRTA and the remaining 82.0% had secondary dRTA.

150

151   The average age of patients with dRTA in 2017 was 53 years (inherited dRTA: 26 years; acquired dRTA:

152   61 years). Approximately 60% of them were women (inherited dRTA: 52%; acquired dRTA: 62%). There

153   were more female (58.0%) patients with dRTA than male patients (42.0%), and no significant differences

154   were observed in the age at diagnosis between genders.

155

156   Of the diagnosed patients in the database (n=216), 55 were recorded as deceased prior to 2017. The mean

157   age of death (SD) was 72 years (13.4 years). There is insufficient data to draw any firm conclusions

158   regarding evidence of excess mortality in this renal condition.

159

160   <u>Associated conditions (renal and non-renal conditions)</u>

161   In primary care settings, nephrocalcinosis (14.4%) and renal stones/calculus (16.7%) were the most

162   commonly coded renal conditions in patients with dRTA. Anemia (19.4%), type II diabetes (19.0%), hearing

163   problems (19.9%), and autoimmune diseases (8.8%), including Sjögren's (Sicca) syndrome and SLE, were

164   the most common non-renal conditions associated with RTA and dRTA (Table 1). Among the 19 patients

165   with autoimmune disease, 18 were diagnosed with dRTA after the age of 20, one was diagnosed at age of

166   12 years.

167

168   The most frequent causes of hospitalization in patients with a coded diagnosis of dRTA were related to

169   kidney or urinary conditions, with the number one cause of hospitalization reported to be disorders

170   secondary to impaired renal tubular function. Similarly, with outpatient visits, patients with dRTA experience

171   a wide range of associated conditions and complications, with hypertension, type II diabetes, hypokalemia,

172   anemia, non-infective gastroenteritis, and colitis being the most frequent causes of hospitalization, aside

173   from renal-related complications.

174

175 <u>Treatments</u>

176 Patient-level data showed that most patients with dRTA are treated with sodium bicarbonate, potassium

177 citrate/citric acid monohydrate, prior to and following diagnosis. Following diagnosis, a substantial increase

178 in mean number of prescriptions were seen for sodium bicarbonate (1.32 vs 5.40 per patient per year;

179 p<0.001). A slight but not significant increase was seen for potassium citrate/citric acid monohydrate (1.93

180 vs 2.45 per patient per year, p=0.657).

181

182 Other than treatment prescribed for renal conditions, the most frequently prescribed drugs in patients with

183 dRTA were found to be for pain/fever management (mostly paracetamol), infections (antibiotics), and

184 anaemia. At the point of diagnosis, prescriptions for non-renal related complications were given in less than

185 5% of patients;

186

187 **Prevalence and extrapolation**

188

189 <u>2017 period-prevalence</u>

190

191 Of the diagnosed patients in the database (n=216), 113 patients were still alive and registered in the CPRD

192 GOLD database in 2017. The algorithm explained in the Methods section (Fig. 1), identified 447 additional

193 patients with suspected dRTA, 280 of which were alive and registered in CPRD GOLD in 2017.

194

195 The resulting prevalence of patients with dRTA in the CPRD database was estimated between 0.46

196 (diagnosed) and 1.60 (diagnosed and suspected) per 10,000 people.

197

198 The number of patients with presumed primary dRTA was between 25 (diagnosed) and 30 (diagnosed +

199 identified through the algorithm) (**Error! Reference source not found.**2) and the number of patients with

200 presumed acquired dRTA was between 88 (diagnosed) and 363 (diagnosed + identified through the

201 algorithm) (Table 2). Among the patients identified with dRTA in 2017, the proportion of cases with

202 presumed primary etiology was 22.1% when considering diagnosed patients only, and 7.6% when

203 considering diagnosed and suspected patients.

204

205 <u>Extrapolation to EU5 countries</u>

206 By applying the prevalence figures calculated in different age and gender groups in the CPRD GOLD

207 database to the European Union Five (EU5) populations, it was estimated that there were 2,989 to 10,427

208 people in the UK with dRTA [0.45 to 1.58 /10,000 person-year], 3,074 to 11,083 people in France [0.46 to

209 1.66 /10,000 person-year], 4,021 to 14,867 people in Germany [0.49 to 1.80 /10,000 person-year], 3,031

210 to 11,230 people in Italy [0.50 to 1.85 /10,000 person-year], and 2,042 to 7,371 people in Spain [0.48 to

211 1.75 /10,000 person-year] in 2017 (Table 3).

212

Discussion

213 **Discussion**

214 There are no epidemiological studies of dRTA in Western countries. Understanding the prevalence and

215 treatment prescribed for dRTA should provide a better understanding of the unmet need and the challenges

216 that patients and physicians face in terms of diagnosis and optimal treatment.

217

218 Our results identified 216 patients with a coded diagnosis of dRTA/RTA in the primary care CPRD GOLD

219 database.

220

221 Because of suspected misclassification/miscoding of dRTA [15], we used our clinical experience to develop

222 an algorithm to identify patients with clinical features of dRTA, yet were not coded as such. This algorithm

223 (Fig. 1) suggested that roughly 2/3 of patients with clinical features of dRTA did not have RTA or dRTA

224 diagnosis codes in CPRD, reinforcing the idea that many patients with dRTA are miscoded or unreported

225 to the GPs. Importantly, these potentially undiagnosed patients were mostly adults (?%), suggesting that

226 especially secondary forms of dRTA are not recognized.

227

228 Main comorbidities found in patients with a coded diagnosis of dRTA included autoimmune diseases and

229 nephrocalcinosis, thus confirming the predominance of these specific comorbidities and the relevance of

230   including them in our algorithm. However, nephrocalcinosis is likely under-reported in CPRD as suggested

231   by the literature: A recent survey conducted on 340 patients through European professional organizations

232   indicated that 88% patients with primary dRTA had nephrocalcinosis [14]. This contrasts with only 23.1%

233   (9 out of 38) of patients with coded dRTA aged less than 20 years old in our study. Similarly, *Both et al.*

234   reported in 2014 that 56 % of patients with dRTA, primary or secondary, had nephrocalcinosis [11].

235   Assuming 88% of patients with primary dRTA have nephrocalcinosis [14] and 18% of dRTA are primary

236   dRTA, we expect that roughly half of patients with secondary dRTA have nephrocalcinosis. Yet, it was only

237   12% in patients with a coded dRTA aged 20 years and older in our study. These relatively low proportions

238   can be explained by a failure of patients or specialists to communicate the results of the series of tests

239   confirming the diagnosis of nephrocalcinosis, an inclination of the GPs to record the underlying diseases,

240   i.e. dRTA, and report nephrocalcinosis only if long-term symptoms such as (chronic) kidney failure are

241   observed.

242   This deficiency in coding nephrocalcinosis may have impaired the ability of our algorithm (which includes

243   nephrocalcinosis as an identifying feature) to capture uncoded primary dRTA cases. Yet, for secondary

244   dRTA, the same proportion of cases with nephrocalcinosis (12%) was retrieved in both the coded and

245   suspected group, suggesting that the use of another identifying feature in the algorithm, the presence of a

246   specific underlying conditions such as auto-immune diseases, was sufficient to capture omitted patients

247   with secondary dRTA.

248   Despite the uncertainty surrounding our algorithm, our estimated prevalence rate in this study (between

249   0.46 and 1.60) is consistent with published estimates, which are between 0.03 and 2.1 per 10,000

250   individuals. [15, 16]. This suggests our assumption that many patients with dRTA may be miscoded and

251   that we can identify these by the classical clinical characteristics of dRTA is valid.

252

253   **Strengths and limitations**

254   This study relied upon clinical expertise to define algorithms for identification of patients with dRTA, as well

255   as on the accuracy of the CPRD GOLD and HES databases.

256

257 There were several limitations to this analysis. Firstly, data from specialists in outpatient settings

258 (diagnoses, prescribed medications) were not perfectly recorded by the GPs in CPRD GOLD. Failure to

259 document identifying features of dRTA used in the algorithm could thus lead to an underestimation of the

260 true prevalence. Secondly, only a subset of the total number of patients identified in the CPRD GOLD

261 database was eligible for CPRD-linked HES data, and therefore, secondary care health records were not

262 available for a proportion of the population. Where HES data was available, it provided a comprehensive

263 range of variables, but did not provide information or data on the medication given to patients during

264 hospitalization, again limiting the ability of our algorithm to detect uncoded patients with dRTA.

265

266 There was also a very low number of patients' test results recorded, of which none contained primary test

267 results and thus no conclusions could be made on the existence or absence of inherited or acquired dRTA

268 using the information recorded in CPRD GOLD.

269

270 The mean age of patients with primary dRTA in 2017 was relatively low (26 years old). This could reflect

271 potential limits of our algorithm to detect primary dRTA, for instance because of insufficient coding for

272 nephrocalcinosis, as discussed above. It may be also due to the limited period of CPRD data that was

273 analyzed in this study (January 1, 1987 to November 27, 2017). In the early years of CPRD GOLD (1987-

274 2010), the disease was likely not well captured. READ codes for RTA/dRTA were only introduced in 2009

275 and clinical genetic testing for this condition only became available in the UK after 2010. Consequently,

276 information was less exhaustively collected, especially data from secondary care.

277

278 Despite recording guidelines, regular quality control, and validity checks encouraging good recording

279 practice, there is no specific administrative process to ensure that each individual patient event is well

280 recorded. Completeness and quality of records rely on the level of compliance from practices to standard

281 process [17-19]. Identification of patients with dRTA requires documented diagnoses and prescriptions (not

282 available in HES data); therefore, patients with dRTA who did not visit a primary care physician or did not

283 receive any diagnosis of interest may not be captured.

284 Many patients are referred by generalists to nephrology departments with a suspicion of dRTA in order to

285 confirm the diagnosis with more specialized investigations. This study assumed that in the absence of

286 contradicting features (e.g. renal Fanconi syndrome), the diagnosis code of dRTA/RTA was always used

287 correctly by the GP. If there is no possibility to demonstrate the certitude of this claim, it is unlikely primary

288 care providers would record a RTA/dRTA diagnosis in the system prior confirmation from a specialist.

289

290 The lack of specific treatment or comorbidities made the identification of false positive challenging and the

291 clinical review of all records shows some limits. Advanced technics such as machine learning could help

292 refine the adjustment.

293

294 Assuming that all diagnosed cases of RTA recorded in CPRD GOLD are dRTA patients may over-estimate

295 the prevalence of dRTA, as it ignores proximal RTA. This has to be balanced by the limitations induced by

296 the data collection process (e.g. under-reporting of diagnosis from GPs) and the number of patients with a

297 recorded diagnosis of RTA/dRTA to consider a reasonable lower border of the estimate of prevalence.

298

299

300 The extrapolation was performed considering age and gender distributions only and therefore assuming

301 that other potential factors influencing the prevalence of dRTA are similar between the UK and other EU5

302 populations. Data supporting or contradicting this assumption are scarce. A retrospective study showed

303 that non-European ethnicity may be associated with increased prevalence of primary Sjögren's syndrome

304 [20], an underlying condition of acquired dRTA. Lack of completeness and consistency of ethnicity recording

305 in CPRD prevented us from using this factor in the extrapolation.

306

307 **Conclusions**

308 For the first time, the prevalence of dRTA in the UK was estimated using large representative primary and

309 secondary care data. This study suggests that the actual prevalence of dRTA is approximately three times

310 higher than reported or coded in health records and that mostly cases of secondary dRTA are being missed.

311   Given the risk of additional complications, such as bone disease, urolithiasis and progressive CKD, there

312   is a need for patients to be diagnosed correctly and as early as possible to ensure that optimal treatment

313   and disease management is implemented.

314

Methods

315

316   To estimate prevalence, data was obtained from the CPRD GOLD database, between January 1, 1987 and

317   November 27, 2017. The research protocol was approved by the Independent Scientific Advisory

318   Committee ethics committee and access to CPRD GOLD and HES data was granted in June 2018.

319

Data source

320

321   CPRD GOLD is a primary care database containing anonymous patient records from GPs. The CPRD

322   GOLD database includes longitudinal information on diagnoses, symptoms, laboratory tests and

323   prescriptions issued by the GP in addition to information on referrals to specialists. For a subset of patients,

324   CRPD is able to link to the HES, a secondary care database of National Health Service hospitals in England

325   that provides information on outpatient, inpatient, and accident and emergency (A&E) patient data.

326

327   Clinical events in the CPRD GOLD are recorded using the "READ code" clinical coding system. Hospital

328   discharge diagnoses in HES are recorded using the international classification of disease (ICD)–10 clinical

329   coding system. Dates of clinical events are precisely recorded, as well as dates of drug prescription

330   delivered in primary care.

331

332   Despite over-representing certain geographical areas of the UK, the CPRD has been found to be

333   representative of the UK population with regard to sex, age and ethnicity [13]. HES data covers all NHS

334   hospital admissions and care delivered by treatment centers funded by the NHS (including those in the

335   independent sector).

336

337 **Study population**

338 The study population is the total number of living individuals recorded in the CPRD GOLD database (around

339 7% of the UK population).

340

341 **Selection criteria to identify diagnosed and undiagnosed patients with dRTA**

342 In order to identify both diagnosed and undiagnosed patients with dRTA, a two-step approach was followed

343 considering the whole registration period of patients in the database, from point of entry to either the date

344 of registration or the last point of record or death.

345

346 Step 1: Patients with diagnosis

347 Step one consisted of identifying all patients with a coded GP diagnosis of dRTA. Since 2009, two READ

348 codes specific for RTA and dRTA have been introduced into the CPRD system, allowing UK primary care

349 providers to record patients with dRTA with exactitude. Patients with any record of RTA (code K08y400;

350 introduced in February 2009), dRTA (code K08yD00; introduced in March 2013) during the study period

351 were included. Refer to Table 4 for all the diagnostic codes included. There were no restrictions on age or

352 gender; however, the analysis included living patients in the database only, for the purpose of estimating

353 prevalence.

354

355 The ability to identify and record patients with dRTA based on specific READ code makes the CPRD a

356 valuable data source for this study. However, the dRTA population will be likely under-reported in CPRD

357 databases, for several reasons:

358 • Firstly, identification of secondary dRTA remains challenging and can be confounded by conditions

359   with similar symptoms. Due to the non-specific nature of the symptoms of dRTA, a thorough clinical

360   evaluation, a variety of specialized tests, and the help of a nephrologist are usually required to

361   make and confirm the diagnosis

362 • Other barriers to clinical coding in primary care practice were identified and reported, such as the

363   skill level of general practice staff, the time it takes to record the data, the motivation of primary

364   care professionals, and the priority that is placed on clinical coding within the organization [21].

365   •   Finally, the coding systems and terminologies themselves are a limit to identify patients with dRTA.
366   [21] While there is a precise READ code in the primary care data, patients referred to hospitals for
367   dRTA are recorded under the 10th revision of the International Statistical Classification of Diseases
368   and Related Health Problems (ICD-10) N25.8: *'Other disorders resulting from impaired renal*
369   *tubular function',* which is not specific. Therefore, if dTRA is not reported by primary care providers,
370   there is no definite evidence in hospital data that allows for the identification of patients with dRTA.

371

372   <u>Step 2: Suspected dRTA cases</u>

373   In step two, an algorithm was developed to detect patients with potential dRTA that were miscoded,

374   undiagnosed, or unreported (Fig. 1). Inclusion and exclusion criteria specific for patients with dRTA were

375   established based on clinical expertise.

376

377   For the purpose of this study we decided that only the combination of the presence of specific comorbidities

378   (autoimmune disease and selected renal conditions), and the use of specific drugs may identify patients

379   with undiagnosed dRTA with a limited risk of misclassification. Laboratory tests and their results were not

380   considered as they were rarely available, and not consistently employed for all patients, preventing the

381   development of a conclusive classification criterion.

382

383   From the source population in CPRD GOLD, 40,560 patients were identified and extracted, who had at

384   least one diagnosis event recorded for dRTA-associated systemic (Sjögren (Sicca) syndrome, systemic

385   lupus erythematosus) or renal disorders (nephrocalcinosis, obstructive uropathy) within the study period

386   [01/01/1987 – 31/12/2017] and no diagnostic code associated with dRTA or RTA (Read codes for identifying

387   events related to the inclusion conditions are presented in table 4). These patients were all fully registered

388   with their general practitioner (GP) and their records had passed CPRD data quality control checks.

389

390   The algorithm identifies suspected patients with dRTA by retaining patients that have the above mentioned

391   clinical diagnoses: Sjögren (Sicca) syndrome, SLE, nephrocalcinosis or obstructive uropathy and in addition

392   receive alkali supplementation (sodium citrate, sodium bicarbonate or potassium citrate).

393   Receiving alkaline treatment was a pre-requisite to be a suspected patient with dRTA. However, a fraction
394   of patients with dRTA, referred as « incomplete dRTA », may have a defect in renal distal acidification
395   without overt metabolic acidosis. These patients may not systematically receive alkaline treatment and
396   therefore, may not be captured in this algorithm. Patients suspected by our algorithm are only those seeking
397   care and receiving treatment.

398

399   Nephrocalcinosis is likely to capture misdiagnosed patients with primary dRTA, although, patients with
400   secondary dRTA can also have nephrocalcinosis (clinical observations). For patients with SLE or
401   obstructive uropathy, we restricted the criteria further by only considering patients with prescriptions of
402   citrates, as we were concerned that prescription of sodium bicarbonate may reflect acidosis due to
403   advanced CKD rather than dRTA (Fig. 1).
404   CPRD-HES-linked patient referrals to specialist physicians in secondary care were also tracked, presuming
405   that patients with dRTA would have regular visits to a nephrologist, a urologist or both. We identified 375
406   patients with autoimmune diseases and prescriptions of alkali agents (Fig. 1). Among them 166 (44.3%)
407   had no visit to nephrologist and urologist recorded and 175 (46.7%) were not linked to HES data. Given the
408   uncertainty around completeness of information from secondary care providers and the large number of
409   patients without reported contact with nephrologist or urologist who satisfied the other criteria, a criterion
410   based on specialist visit was not considered as a key element of the algorithm and discontinued.

411

412   **Data analysis**
413   The primary outcome of the study was the point prevalence (per 10,000 people) of patients with dRTA at
414   the end of the study period (November 27, 2017) [22]. The analysis reported the number of patients
415   (prevalent cases) with a confirmed diagnosis of dRTA, the number of suspected cases (undiagnosed,
416   unreported or miscoded) and the lower and upper bounds of the prevalence in the CPRD population, where
417   the lower bound is represented by diagnosed patients with dRTA and the upper bound is the number of
418   diagnosed and suspected patients with dRTA.

419

420 Therefore, the lower bound prevalence of dRTA was calculated by dividing the number of living patients

421 with a confirmed diagnosis of dRTA by the total number of living patients recorded in the CPRD GOLD

422 database. The upper bound prevalence of dRTA was calculated by dividing the number of living patients

423 diagnosed with or suspected of having dRTA by the total number of living patients recorded in the CPRD

424 GOLD database. The prevalence rates by age and gender observed in this study were also applied to EU5

425 population structures.

426

427 Secondary outcomes included patient demographics, characteristics, comorbidities and pharmacological

428 treatments in patients with a diagnosis of dRTA. For categorical variables, the study reported the sample

429 size and frequency. For continuous variables, measures of mean and standard deviation (SD) are reported.

430 Outcomes were reported by age groups and type of dRTA (primary or secondary), when relevant. We

431 defined primary dRTA as patients diagnosed before the age of 20 years and secondary dRTA as patients

432 diagnosed after the age of 20 years.

433

434 **List of abbreviations**

435 A&E, Accident and Emergency

436 CPRD, Clinical Practice Research Datalink

437 dRTA, distal renal tubular acidosis

438 GP, general practitioner

439 HES, Hospital Episode Statistics

440 RTA, renal tubular acidosis

441 SD, standard deviation

Declarations

Ethics approval and consent to participate

The research protocol was approved by the Independent Scientific Advisory Committee ethics committee, and access to CPRD GOLD and HES data was granted in June 2018.

Consent for publication

Not applicable

Availability of data and materials

The data that support the findings of this study are available from the CPRD but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of the CPRD.

Competing interests

Funding

Authors' contributions

FB was involved in the design of the study, the data acquisition, the interpretation of the results, and the writing of the manuscript. FG was involved in the statistical analyses and interpretation of the results, and

470   the writing of the manuscript. LR was involved in the design of the study, the interpretation of the results,

471   and the critical revision of the manuscript. CM was involved in the design of the study, the interpretation

472   of the results, and the critical revision of the manuscript. DG was involved in the definition of the algorithm

473   and the critical revision of the manuscript. DB was involved in the definition of the algorithm and the

474   critical revision of the manuscript. All authors read and approved the final manuscript.

475

476   **Acknowledgements**

478

19

References

1. Lim, A.K. and M.J. Choi, *Distal renal tubular acidosis associated with Sjogren syndrome.* Intern Med J, 2013. **43**(12): p. 1330-4.
2. Escobar, L., et al., *Distal renal tubular acidosis: a hereditary disease with an inadequate urinary H(+) excretion.* Nefrologia, 2013. **33**(3): p. 289-96.
3. Basak RC, et al., *Distal Renal Tubular Acidosis, Hypokalemic Paralysis, Nephrocalcinosis, Primary Hypothyroidism, Growth Retardation, Osteomalacia and Osteoporosis Leading to Pathological Fracture: A Case Report.* Oman Medical Journal, 2011. **26**(4): p. 271-274.
4. Domrongkitchaiporn, S., et al., *Bone mineral density and histology in distal renal tubular acidosis.* Kidney Int, 2001. **59**(3): p. 1086-93.
5. Zhang, C., et al., *Clinical evaluation of Chinese patients with primary distal renal tubular acidosis.* Intern Med, 2015. **54**(7): p. 725-30.
6. Fry, A.C. and F.E. Karet, *Inherited renal acidoses.* Physiology (Bethesda), 2007. **22**: p. 202-11.
7. Orphanet. *Rare diseases.* 2018 [cited 2019 16 December]; Available from: http://www.orpha.net/consor/cgi-bin/OC_Exp.php?Expert=18&lng=EN.
8. Sarah S, et al., *Renal tubular acidosis due to Sjorgen's syndrome presenting as hypokalemic quadriparesis: A report of two cases.* Indian J Nephrol, 2015. **25**(6): p. 386-387.
9. Genetics Home Reference. *SLC4A1 – associated distal renal tubular acidosis.* 2018 [cited 2019 17 December]; Available from: https://ghr.nlm.nih.gov/gene/SLC4A1#conditions.
10. Batlle, D. and S.K. Haque, *Genetic causes and mechanisms of distal renal tubular acidosis.* Nephrol Dial Transplant, 2012. **27**(10): p. 3691-704.
11. Both, T., et al., *Everything you need to know about distal renal tubular acidosis in autoimmune disease.* Rheumatol Int, 2014. **34**(8): p. 1037-45.
12. Both, T., et al., *Prevalence of distal renal tubular acidosis in primary Sjogren's syndrome.* Rheumatology (Oxford), 2015. **54**(5): p. 933-9.
13. Herrett, E., et al., *Data Resource Profile: Clinical Practice Research Datalink (CPRD).* Int J Epidemiol, 2015. **44**(3): p. 827-36.
14. Lopez-Garcia, S.C., et al., *Treatment and long-term outcome in primary distal renal tubular acidosis.* Nephrology Dialysis Transplantation, 2019. **34**(6): p. 981-991.
15. Bryant G, L.L., Li-McLeod J, *Estimate of Prevalence of Secondary Distal Renal Tubular Acidosis Among Patients with Sjogren's Syndrome and Systemic Lupus Erythematosus in a US Population with Employer-Sponsored Health Insurance [abstract].* Arthritis Rheumatol, 2019. **71 (suppl 10)**.
16. Silva, C., et al., *PUK20 ESTIMATE OF PREVALENCE OF PRIMARY DISTAL RENAL TUBULAR ACIDOSIS AMONG THE US POPULATION WITH EMPLOYER-SPONSORED HEALTH INSURANCE.* Value in Health, 2019. **22**: p. S384.
17. Ghosh, R.E., et al., *How Clinical Practice Research Datalink data are used to support pharmacovigilance.* Ther Adv Drug Saf, 2019. **10**: p. 2042098619854010.
18. Tate, A.R., et al., *Quality of recording of diabetes in the UK: how does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study using the CPRD database.* BMJ Open, 2017. **7**(1): p. e012905.
19. Khan, N.F., S.E. Harrison, and P.W. Rose, *Validity of diagnostic coding within the General Practice Research Database: a systematic review.* Br J Gen Pract, 2010. **60**(572): p. e128-36.
20. Maldini, C., et al., *Epidemiology of primary Sjogren's syndrome in a French multiracial/multiethnic area.* Arthritis Care Res (Hoboken), 2014. **66**(3): p. 454-63.
21. de Lusignan, S., *The barriers to clinical coding in general practice: a literature review.* Med Inform Internet Med, 2005. **30**(2): p. 89-97.
22. Hunter PR and Risebro H, *Defining the current situation – epidemiology*, in *Valuing Water, Valuing Livelihoods*, Cameron J, et al., Editors. 2011, IWA Publishing: London.

Table 1. Characteristics of patients with a coded diagnosis of dRTA/RTA

	All	dRTA diagnosis recorded before age of 20	dRTA diagnosis recorded after age of 20
Identified diagnosed cases of dRTA/RTA in CPRD	**216**	**38**	**178**
Age at diagnosis [mean (sd)]	46.1 (25.9)	7.1 (7.5)	54.4 (20.2)
Follow-up duration in years [mean (sd)]	22.7 (15.8)	14.5 (9.5)	24.4 (16.3)
Gender (male)	91 (42.1%)	18 (47.4%)	73 (41.1%)
Renal conditions (ever), (n, %)			
Nephrocalcinosis	31 (14.35%)	9 (23.7%)	22 (12.4%)
Renal Stone	12 (5.56%)	1 (2.6%)	11 (6.2%)
Calculus of Kidney	10 (4.63%)	2 (5.3%)	8 (4.5%)
Medullary Sponge Kidney	7 (3.24%)	-	7 (3.9%)
Renal Stone – uric acid	7 (3.24%)	-	7 (3.9%)
Ureteric Stone	4 (1.85%)	2 (5.3%)	2 (1.1%)
Renal calculus	3 (1.39%)	-	3 (1.7%)
Calculus of kidney and ureter	2 (0.93%)	-	2 (1.1%)
Obstructive uropathy	1 (0.46%)	-	1 (0.6%)
Calculus of kidney with calculus of ureter	1 (0.46%)	-	1 (0.6%)
Calculus in urethra	1 (0.46%)	1 (2.6%)	-
Staghorn calculus	1 (0.46%)	-	1 (0.6%)
Nephropathy, unspecified	1 (0.46%)	-	1 (0.6%)
Main recorded non-renal comorbidities (ever), most frequent (n, %)			
Anaemia unspecified	42 (19.44%)	3 (7.9%)	39 (21.9%)
Type II diabetes mellitus	41 (18.98%)	-	41 (23.0%)
Hearing Difficulty	19 (8.80%)	5 (13.2%)	14 (7.9%)
Sicca (Sjogren's) syndrome	15 (6.94%)	1 (2.6%)	14 (7.9%)
Hearing Loss	14 (6.48%)	4 (10.5%)	10 (5.6%)
Sensorineural hearing loss	9 (4.17%)	2 (5.3%)	7 (3.9%)
Systemic Lupus Erythematosus	6 (2.78%)	-	6 (3.4%)
Hyperparathyroidism	6 (2.78%)	-	6 (3.4%)
Rheumatoid Arthritis	6 (2.78%)	1 (2.6%)	5 (2.8%)
Hyperthyroidism	3 (1.39%)	-	3 (1.7%)
Infective Hepatitis	2 (0.93%)	-	2 (1.1%)

Lupus Erythematosus	2 (0.93%)	-	2 (1.1%)
Hearing impairement	1 (0.46%)	-	1 (0.6%)
Lupus Nephritis	1 (0.46%)	-	1 (0.6%)
Autoimmune disease (n, %)	**19 (8.8%)**	**1 (2.6%)**	**18 (10.1%)**
Autoimmune disease or nephrocalcinosis or obstructive uropathy (n, %)	**46 (21.3%)**	**9 (23.7%)**	**37 (20.8%)**

530

531

Table 2. Number of diagnosed and suspected cases of dRTA, inherited or acquired identified in the CPRD GOLD database

534

	Number of cases of dRTA throughout 1987–2017			Prevalent cases of dRTA in 2017		
	All	**Before 20 years**	**After 20 years**	**All**	**Before 20 years**	**After 20 years**
Patients identified based on a coded diagnosis of dRTA/RTA in CPRD GOLD	**216**	**38**	**178**	**113**	**25**	**88**
Suspected cases of dRTA in CPRD GOLD	**447**	**8**	**439**	**280**	**5**	**275**
Among them:						
(A) With auto-immune diseases + alkali agents[1]	375	0	375	240	0	240
(B) With nephrocalcinosis + alkali agents	55	8	47	34	5	29
(C) Wiith obstructive uropathy + alkali agents	17	0	17	6	0	6
Diagnosed cases and suspected cases	**663**	**47**	**616**	**393**	**30**	**363**
CPRD, Clinical Practice Research Datalink; dRTA, distal renal tubular acidosis; [1]Excluding patients with lupus who were prescribed sodium bicarbonate only. *Note: In absence of codes diagnosis for dRTA/RTA and assuming patients were not reported or miscoded, the potential diagnosis date was approximated using the first prescription of sb, pc or sc.; Only eight suspected patients presented entered in the definition before the age of 20 years old.* (A), (B), (C): correspondence with figure 1.						

535

536

537 **Table 3. Extrapolation to EU5 countries (number of cases and prevalence)**

		dRTA (confirmed)	dRTA (confirmed and suspected)
In the CPRD database	Prevalence (/10,000 people)	0.46	1.60
Extrapolation to EU5 countries*			
UK	Number of cases with dRTA	2989	10427
	Prevalence (/10,000 people)	0.45	1.58
France	Number of cases with dRTA	3074	11083
	Prevalence (/10,000 people)	0.46	1.66
Germany	Number of cases with dRTA	4021	14867
	Prevalence (/10,000 people)	0.49	1.80
Italy	Number of cases with dRTA	3031	11230
	Prevalence (/10,000 people)	0.50	1.85
Spain	Number of cases with dRTA	2042	7371
	Prevalence (/10,000 people)	0.48	1.75
dRTA, distal renal tubular acidosis; EU5, European Union Five.			
*Number of cases and prevalence estimated by applying the prevalence figures calculated in different age and gender groups in the CPRD to the EU5 populations.			

538

539

540 **Table 4. Clinical codes used to identify patients diagnosed, or potentially undiagnosed, with dRTA**

541 **in CPRD data**

| Patients diagnosed with dRTA		
Condition	**Type of field**	**Code**
RTA	medcode	5072
dRTA	medcode	105829
Potential undiagnosed patients		
Condition/specialist	**Type of field**	**Code**
Visits to specialist		
Nephrologist	mainspef	361
Urologist	mainspef	101
Conditions		
Stones	medcode	2258, 4928, 10282, 9162, 6048, 9950, 1858
Sicca	medcode	2360
Lupus	medcode	47672, 58706, 29519, 42719, 36942, 7871, 22205
Autoimmune hepatitis	medcode	18652
Hearing loss	medcode	5967, 9830, 536, 96245, 18008
Nephrocalcinosis	medcode	8690, 56258, 111458
Obstructive uropathy	medcode	12095
Hypertension	medcode	799, 3425, 4444, 13186, 19070, 3712, 10818, 13239, 6378
Chronic kidney disease	medcode	95406, 95405, 12479 105151, 12585, 104963, 95122, 95508
Dialysis	medcode	20073, 11773, 2996, 2994, 20196
Renal transplant	medcode	18774, 17253
Treatments		
Sodium bicarbonate	Drugsubstance*	SODIUM BICARBONATE
Sodium citrate	Drugsubstance*	SODIUM CITRATE
Potassium citrate	Drugsubstance*	POTASSIUM CITRATE
Everolimus	Drugsubstance*	EVEROLIMUS
Sirolimus	Drugsubstance*	SIROLIMUS
Mycophenolate mofetil	Drugsubstance*	MYCOPHENOLATE MOFETIL
Ciclosporin	Drugsubstance*	CICLOSPORIN, CYCLOSPORIN
*Transformed in capital letters. CPRD, Clinical Practice Research Datalink; dRTA, distal renal tubular acidosis; RTA; renal tubular acidosis.		

542

543

Figure 1. Algorithm developed to identify potential patients with dRTA

545
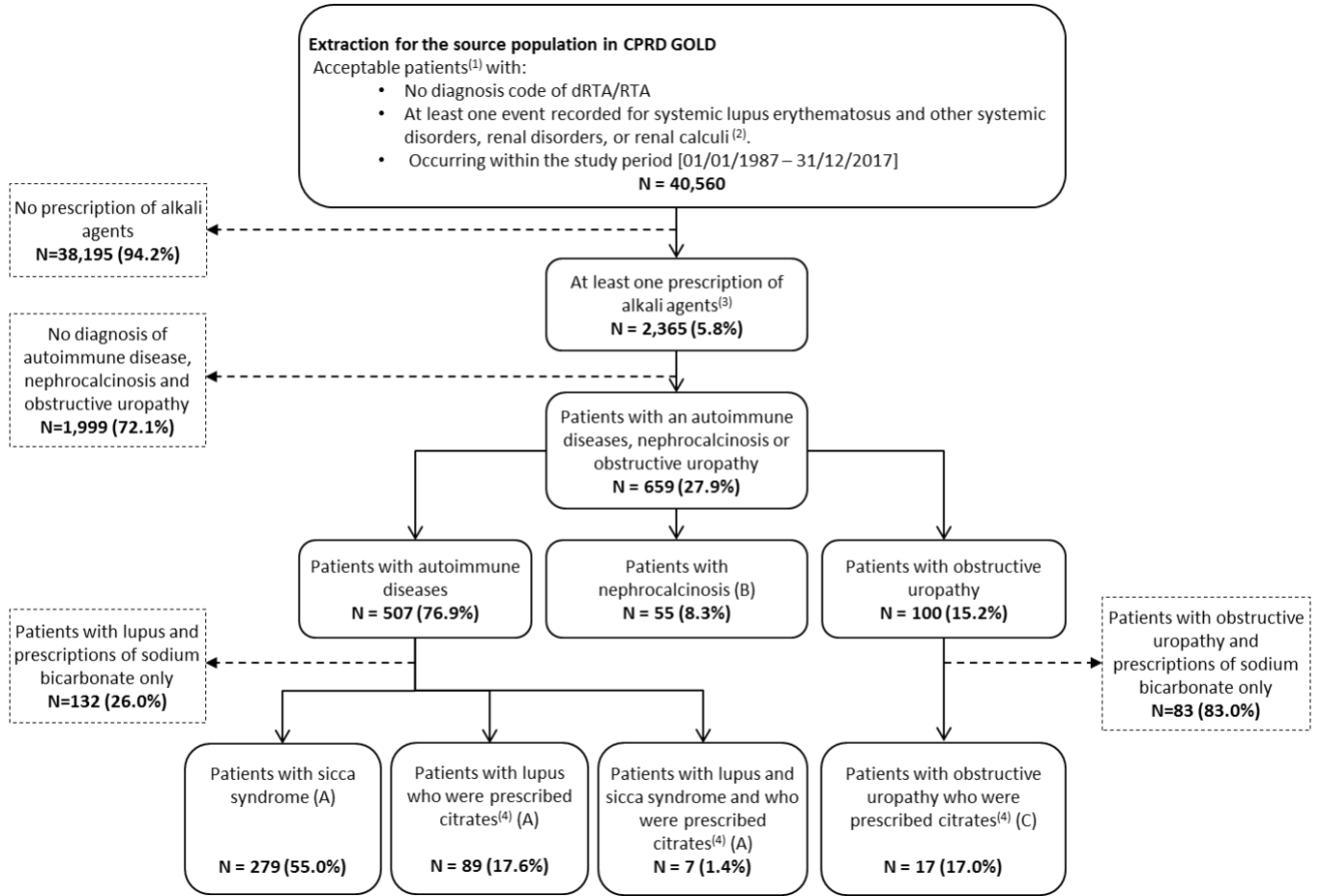
546

547

548 dRTA, distal renal tubular acidosis; HES, Hospital Episode Statistics; RTA, renal tubular acidosis;
549 (1) Patients with an up-to-standard (UTS) follow-up i.e. with a follow-up period of good quality data from the practice,
550 as defined by CPRD; (2) See medical codes in Table 4; (3) sodium bicarbonate, potassium citrate or sodium citrate
551 (4) potassium citrate or sodium citrate or both; patients with only a prescriptions of sodium bicarbonate only were
552 removed.
553 *Notes: A total of 447 suspected patients with dRTA were identified [=375 (A) + 55(B) + 17(C)]; Three patients with*
554 *autoimmune diseases had nephrocalcinosis.*
555 (A), (B), (C): correspondence with Table 1

556