Table S1. List of keywords and number of matches found in text of clinical notes (see attached excel file).

TABLE S2. List of 226 keywords used to identify individuals with rare epilepsies in electronic health record data

| **Rare Epilepsy** | **Keywords** |
|---|---|
| Aicardi syndrome | Aicardi's syndrome, Aicardi, Aicardi's, retinal lacunae, agenesis of corpus callosum, absent corpus callosum, porencephalic cysts, coloboma, retinal lesions, lacunae |
| Alpers disease | 15q24, Alpers Disease, Alpers syndrome, Alpers-Huttenlocher, Alpers-Huttenlocher syndrome, Alpers |
| Angelman syndrome | uniparental disomy, UBE3A, happy puppet, Angelman's syndrome, 15q11, Angelman syndrome |
| CDKL5 | X-linked infantile spasm, STK9, atypical rett syndrome, atypical rett, CDKL5 |
| Dravet syndrome | GABRD, GABRG2, GEFS+, genetic epilepsy with febrile seizures plus, PCDH19, protocadherin, severe myoclonic epilepsy of infancy, SIEE, SMEB, SMEI, Dravet |
| Dup15q syndrome | idic(15), idic 15, dup15q syndrome, dup15q, 15q11 duplication, 15q11 microduplication, duplication of 15q, duplication 15q, 15q11.2 |
| Early infantile developmental and epileptic encephalopathy | EIEE, early infantile epileptic encephalopathy, infantile epileptic encephalopathy, Ohtahara, suppression burst, suppression-burst |
| EE-SWAS/ESES | electrographic status epilepticus in sleep, electrical status epilepticus of sleep, ESES index, continuous slow spike and wave of sleep, spike-index, spike index, CSWS, acquired epilep, autistic epilep, continuous spike-wave in sleep, continuous spike and wave in sleep, 16p13.2, electrographic status epilepticus of sleep, epileptic aphasia, ESES, ESES with language regression, GRIN2A, LK syndrome, LKS |
| EMAS | Doose, SCN1B, myoclonic-astatic epilepsy, myoclonic atonic epilepsy, GABRG2, EMAS, Doose syndrome, atonic, astatic, myoclonic atonic, myoclonic astatic |
| Epilepsy in infancy with migrating focal seizures | KCNT1, malignant migrating partial seizures in infancy, migrating partial epilepsy of infancy, MMPSI, SLC25A22, TBC1D24, migrating partial |
| Fragile X syndrome | FMR1, frax, macroorchidism |
| Glut1 deficiency | glut1, glut 1, dystonia 18, dystonia 9, glucose transporter deficiency, glucose transporter type 1 deficiency, glut-1 deficiency syndrome, SLC2A1, SLC2A1 mutation |
| Holoprosencephaly | lobar, semilobar, HPE, sonic hedgehog, SHH |
| Hypothalamic hamartoma with seizures | Pallister-Hall, hypothalamic hamartoma, gelastic epilepsy, gelastic seizures, hypothalamic lesion |
| Infantile spasms | West syndrome, infantile myoclonic epilepsy, jacknife, hypsarrhythmia, salaam, hyps, infantile spasms |

| | |
|---|---|
| KCNQ2 related epilepsy | potassium channelopathy, potassium channel mutations, fifth day fits, familial neonatal seizures, EBN, early onset epileptic encephalopathy, early infantile epileptic encephalopathy, BFNC, benign neonatal seizures, benign neonatal epilepsy, benign familial neonatal seizures, benign familial neonatal epilepsy, benign familial neonatal convulsions, KCNQ2 |
| Lennox-Gastaut syndrome | multiple seizure types, lennox syndrome, lennox gastaut, slow spike-wave, slow spike and wave |
| Myoclonic epilepsy with ragged red fibers | MERRF, MERF, MERFF, MERRF syndrome, RRF, MTTK, MTTL1, MTTH, MTTS1, MTND5 |
| Neuronal ceroid lipofuscinosis | NCL, Batten's disease, Kufs disease, Battens, ceroid, Batten disease, CLN, LINCL |
| PCDH19 | EFMR, cluster disorder, PCDH19 mutation, protocadherin, PCDH19 |
| Phelan-McDermid syndrome | SHANK3, Phelan-McDermid syndrome, 22q13 deletion, Phelan-McDermid, Phelan, 22q13 |
| Prader Willi syndrome | PWS, SNRPN, uniparental disomy, 15q11, Prader Willi syndrome, Prader Willi, Prader-Willi, Prader-Wili, Prader Wili, Prader |
| Rasmussen syndrome | Rasmussen's encephalitis, chronic focal encephalitis, Rasmussen, Rasmussen's syndrome |
| Rett syndrome | Rett, RTS, Rett's disease, Rett's, Rett syndrome, Rett synd, MECP2, FOXG1, hand wringing, Rolando |
| Ring Chromosome 14 | r14, ring 14, 14q32, ring chromosome 14 |
| Ring Chromosome 20 | ring chromosome 20 syndrome, ring 20, r(20), ring chromosome 20 |
| SCN2A | SCN2A, SCN2A mutations, benign familial neonatal-infantile seizures |
| SCN8A | CIAT, early infantile epileptic encephalopathy, sodium channelopathy, SCN8A |
| SLC13a5 | SLC3a5 |
| Sturge-Weber syndrome | SWS, Sturge Weber syndrome, Sturge-Weber syndrome, Sturge Weber, Sturge-Weber, port-wine stain, port wine stain |
| SYNGAP | RASA1 |
| Tuberous sclerosis complex | TSC, tuberous sclerosis complex, TSC1, TSC2, multifocal micronodular pneumocyte hyperplasia, radial migration lines |
| Unverricht-Lundborg Disease | Baltic myoclonus, EPM1, CSTB, progressive myoclonic epilepsy, lysosome |

Abbreviations: epileptic encephalopathy with spike-and-wave activation in sleep (EE-SWAS) and/or electrical status epilepticus in sleep (ESES), epilepsy with myoclonic atonic seizures (EMAS).

TABLE S3. Regular expressions for creating rare epilepsy cohorts using electronic health record data

| Rare Epilepsy | Regular Expression |
|---|---|
| Aicardi syndrome | "aicardi\|retinal lacunae" |
| Alpers disease | "alpers disease\|alpers syndrome\|alpers-huttenlocher" |
| Angelman syndrome | "angelman['s]* syndrome\|15q11" |
| CDKL5 | "x-linked infantile spasm\|cdkl5" |
| EMAS | "doose\|myoclonic( \|-)astatic\|myoclonic atonic epilepsy\|epilepsy with myoclonic atonic seizures\|emas" |
| Dravet syndrome | "gabrd\|severe myoclonic epilepsy of infancy\|smeb\|dravet" |
| Dup15q syndrome | "idic 15\|dup15q\|15q11 duplication\|15q11 microduplication\|duplication of 15q\|duplication 15q\|15q11.2" |
| Glut1 deficiency | "dystonia 9\|glucose transporter type 1 deficiency\|glut[- ]*1\|slc2a1" |
| Holoprosencephaly | "semilobar\|holoprosencephaly" |
| Hypothalamic hamartoma with seizures | "hypothalamic hamartoma\|gelastic epilepsy\|gelastic seizures" |
| Infantile spasms | "jacknife\|hyps\|infantile spasm" |
| KCNQ2 related epilepsy | "fifth day fits\|familial neonatal seizures\|early infantile epileptic encephalopathy\|benign neonatal epilepsy\|kcnq2" |
| EE-SWAS/ESES | "electrographic status epilepticus (in\|of) sleep\|electrical status epilepticus of sleep\|eses index\|continuous slow spike and wave of sleep\|spike-index\|acquired epilep\|continuous spike-wave in sleep\|epileptic aphasia\|eses with language regression\|lk syndrome\|lks\|epileptic encephalopathy with spike[- ]*and[- ]*wave activation in sleep\|ee[- ]*swas" |
| Lennox-Gastaut syndrome | "lennox syndrome\|lennox gastaut\|slow spike-wave\|slow spike and wave" |
| Epilepsy in infancy with migrating focal seizures | "kcnt1\|mmpsi\|migrating partial\|epilepsy in infancy with migrating focal seizure\|eimfs" |
| Myoclonic epilepsy with ragged red fibers | "merrf" |
| Neuronal ceroid lipofuscinosis | "batten['s]* disease\|[li]*ncl" |
| Early infantile developmental and epileptic encephalopathy | "ohtahara\|early infantile [developmental ]*epileptic encephalopathy\|eid*ee" |
| PCDH19 | "efmr\|pcdh19" |
| Phelan-McDermid syndrome | "phelan-mcdermid\|22q13" |
| Prader Willi syndrome | "prader willi" |
| Rasmussen syndrome | "rasmussen" |
| Rett syndrome | "rett's\|rett syndrome" |
| Ring Chromosome 14 | "ring 14" |
| Ring Chromosome 20 | "ring 20\|ring chromosome 20" |
| SCN2A | "scn2a" |
| SCN8A | "scn8a" |

| | |
|---|---|
| Sturge-Weber syndrome | "sturge weber\|sturge-weber" |
| Tuberous sclerosis complex | "tsc\|tuberous sclerosis\|multifocal micronodular pneumocyte hyperplasia\|radial migration lines" |

Abbreviations: epileptic encephalopathy with spike-and-wave activation in sleep (EE-SWAS) and/or electrical status epilepticus in sleep (ESES), epilepsy with myoclonic atonic seizures (EMAS).