Received: 12 June 2023 | Revised: 25 August 2023 | Accepted: 28 August 2023

DOI: 10.1111/epi.17766

Epilepsia™

RESEARCH ARTICLE

# Genetic influences on epilepsy outcomes: A whole-exome sequencing and health care records data linkage study

**Beata Fonferko-Shadrach[1] | Arron S. Lacey[1] | Huw Strafford[1] | Carys Jones[1] | Mark Baker[2] | Robert Powell[1,2] | Ashley Akbari[1] | Ronan A. Lyons[1] | David Ford[1] | Simon Thompson[1] | Kerina H. Jones[1] | Seo-Kyung Chung[1,3,4] | William O. Pickrell[1,2] | Mark I. Rees[1,5]**

[1]Faculty of Medicine, Health, & Life Science, Swansea University Medical School, Swansea, UK

[2]Swansea Bay University Health Board, Swansea, UK

[3]Brain & Mind Centre, University of Sydney, Camperdown, New South Wales, Australia

[4]Kids Research, Children's Hospital at Westmead, Sydney, New South Wales, Australia

[5]Faculty of Medicine & Health, University of Sydney, Camperdown, New South Wales, Australia

**Correspondence**
Mark I. Rees, Faculty of Medicine & Health, University of Sydney, L2.05 the Quadrangle, Camperdown, NSW 2006, Australia.
Email: mark.rees@sydney.edu.au

## Abstract

**Objective:** This study was undertaken to develop a novel pathway linking genetic data with routinely collected data for people with epilepsy, and to analyze the influence of rare, deleterious genetic variants on epilepsy outcomes.

**Methods:** We linked whole-exome sequencing (WES) data with routinely collected primary and secondary care data and natural language processing (NLP)-derived seizure frequency information for people with epilepsy within the Secure Anonymised Information Linkage Databank. The study participants were adults who had consented to participate in the Swansea Neurology Biobank, Wales, between 2016 and 2018. DNA sequencing was carried out as part of the Epi25 collaboration. For each individual, we calculated the total number and cumulative burden of rare and predicted deleterious genetic variants and the total of rare and deleterious variants in epilepsy and drug metabolism genes. We compared these measures with the following outcomes: (1) no unscheduled hospital admissions versus unscheduled admissions for epilepsy, (2) antiseizure medication (ASM) monotherapy versus polytherapy, and (3) at least 1 year of seizure freedom versus <1 year of seizure freedom.

**Results:** We linked genetic data for 107 individuals with epilepsy (52% female) to electronic health records. Twenty-six percent had unscheduled hospital admissions, and 70% were prescribed ASM polytherapy. Seizure frequency information was linked for 100 individuals, and 10 were seizure-free. There was no significant difference between the outcome groups in terms of the exome-wide and gene-based burden of rare and deleterious genetic variants.

**Significance:** We successfully uploaded, annotated, and linked genetic sequence data and NLP-derived seizure frequency data to anonymized health care records in this proof-of-concept study. We did not detect a genetic influence on real-world epilepsy outcomes, but our study was limited by a small sample size. Future

studies will require larger (WES) data to establish genetic variant contribution to epilepsy outcomes.

**KEYWORDS**

antiseizure medication, genetic variant burden, natural language processing, seizure frequency, unscheduled admission

## 1 | INTRODUCTION

Epilepsy is one of the most common neurological conditions, affecting approximately 50 million people worldwide.[1] People with epilepsy face significant physical health, mental health, and social problems in addition to the ongoing risk of seizures.[2] It is established that genetics plays a key role in epilepsy etiology, and significant advances have been made in understanding its underlying genetic architecture.[3] For example, we now know that both single-gene/polygenic rare and common genetic variants contribute to the risk of developing epilepsy.[4,5]

Understanding the cause of a person's epilepsy or correctly identifying their epilepsy syndrome will guide clinical management and prognostication. However, two people with the same epilepsy syndrome or cause for their epilepsy may have differing outcomes, for example, different responses to medication, seizure frequencies, or risks of epilepsy-related death. Many factors may influence these outcomes, with genotype playing an important part. Pharmacogenomic studies have identified genes associated with poor response to antiseizure medications (ASMs), such as mutations in genes encoding *CYP* enzymes, transporter genes, and genes associated with seizures that are also linked to pharmacoresistance, such as *SCN1A*.[6,7]

Sudden unexplained death in epilepsy (SUDEP), the most severe epilepsy outcome, is associated with an increased polygenic burden and a greater presence of potentially deleterious variants. However, no single locus has an association with SUDEP.[8] Can similar genome burden associations be found in common epilepsies with more drug-resistant seizures or real-world outcomes like epilepsy-related hospital admissions?

Using routinely collected data for epilepsy research has many advantages, including less recruitment bias (whole populations can be studied), the relative ease of using data from different health sources, and that the data have already been collected.[9–11] The use of genetic sequencing in routine clinical practice and clinical trials is increasing, with most patients in the near future being likely to have some form of genetic sequencing as part of their routine care.[12,13] There is an exciting opportunity, therefore, to use this routinely collected genetic data to study the effect of epilepsy genotypes on clinically relevant epilepsy outcomes on a population level.

**Key Points**

- We investigated the influence of rare, deleterious genetic variants on real-world epilepsy outcomes
- We linked WES data to routinely collected health care data on unscheduled hospital admissions, ASM therapy, and seizure frequency
- Gene variant annotation, data linkage, and analysis were performed within the Secure Anonymised Information Linkage Databank's trusted research environment
- No association between the genetic burden and unscheduled hospital admissions, ASM polytherapy, and increased seizure frequency was seen
- Future studies will require larger WES data to establish genetic variant contribution to epilepsy outcomes

In this capability pathfinder project, we linked genetic data with routinely collected health data, including that derived from clinic letters using a natural language processing (NLP) application, within the world-leading Secure Anonymised Information Linkage (SAIL) Databank, to study genetic influences on epilepsy outcomes in Wales. Our objectives were to (1) develop a novel pathway to link genetic data with anonymized routinely collected data for people with epilepsy; (2) develop methods to analyze genetic sequencing data within a trusted research environment for anonymized routinely collected data; and (3) use successful completion of (1) and (2) to study the influence of rare, deleterious genetic variants on the risk of ASM polytherapy, unscheduled hospital admissions for epilepsy, and increased seizure frequency.

## 2 | MATERIALS AND METHODS

### 2.1 | Participants

We selected individuals from within the Swansea Neurology Biobank (SNB). SNB participants were recruited from

National Health Service neurology clinics with sex, ethnicity, and detailed epilepsy and medical history recorded from interviews and medical notes review. All individuals in this study had given informed written consent to share their genetic and clinical data anonymously with the SAIL Databank.

Whole-exome sequencing (WES; Illumina HiSeqX platform) was performed at the Broad Institute as part of the Epi25 collaborative. Epi25 is a global collaborative (2014 to present), with the aim of exome sequencing up to 25 000 patients with epilepsy across the world (http://epi-25.org/). SNB has participated in the project from its onset, and exomes were repatriated by downloading data from the Broad Institute central core. Our cohort for this study was submitted to Years 1–3 of Epi25 and derived from patients who donated samples primarily between 2016 and 2018. Epilepsy diagnosis was verified for all participants prior to Epi25 submission and categorized into focal or generalized onset.[14] Detailed phenotypic data from the SNB database and Epi25 WES data (Variant Call Format, VCF v4.2 files) for 111 individuals were used in this project.

## 2.2 | Routinely collected data and data linkage

The SAIL Databank is a privacy-protecting trusted research environment holding a repository of routinely collected anonymized health, social, and administrative data for the Welsh population, and provides infrastructure for linkage and access to project-specific data.[15] Person-identifiable information is separated from nonidentifiable data and is encrypted and substituted with a unique anonymous field used for data linkage.[16] Primary care data in the SAIL Databank are held in the Welsh Longitudinal General Practice dataset and contain dated and coded clinical events from general practitioners (GPs) for approximately 85% of the Welsh population. The dataset includes symptoms, diagnoses, and treatments recorded during primary health care consultations and coded using the Read Code system.[17] Secondary care data in the SAIL Databank are held in the Patient Episode Database for Wales and contain diagnostic and treatment information relating to inpatient/day case episodes for Welsh patients, using the International Classification of Disease version 10 (ICD-10) and Office of Population Census and Surveys version 4 diagnostic and procedural codes.

We uploaded phenotypic and demographic data from the SNB together with WES variant files (VCF 4.2 files) for the study cohort into the SAIL Databank using the established split-file approach.[18] We linked these data to the primary and secondary care datasets within the SAIL Databank. We used the Extraction of Epilepsy Clinical Text (ExECT) NLP pipeline to extract seizure frequency information from epilepsy clinic letters from the medical notes of the individuals within the study group.[19] In a representative sample of letters, the ExECT V2 pipeline extracted seizure frequency with a precision, recall, and F1 score of 74%, 64%, and 69% per item and 92%, 83%, and 88% per letter, respectively.[20] We derived seizure severity scores (SSSs) from the quantifiable expressions of seizure frequency.[21] These expressions were converted to daily rates; for time periods expressed as a range (e.g., 3–4 weeks), the lower number was selected as the denominator; when seizures were expressed as a range (e.g., 5–10 seizures), the higher number was selected as the numerator, so as to avoid underestimating the reported rate. SSSs range from 1 (seizure-free for ≥2 years) to 7 (daily seizures). We uploaded this information into the SAIL Databank, linking it to the primary care and WES data (Figure 1).

## 2.3 | Epilepsy cohort definitions and outcomes

We used primary care Read Codes to identify epilepsy diagnoses (Read v2 "F25"), ASM prescriptions (Read v2 "dn%" and "do%"), and dates within the primary care records dataset.[22] The earliest date of ASM prescription or epilepsy diagnosis was set as the beginning of the study window for each patient (not earlier than January 1, 2000) and the latest as the end (not later than December 31, 2019). To identify unscheduled hospital admissions for epilepsy, we used admission method codes "21" and "29," in combination with the ICD-10 G40 diagnosis.

We created three outcome groups for analysis: (1) study participants without unscheduled hospital admission for epilepsy as a primary cause of admission compared to those with at least one such admission, (2) Study participants prescribed one ASM (monotherapy) versus those prescribed two or more ASMs for at least 6 months (polytherapy), and (3) Individuals who were seizure-free for 1 year or more (SSS = 1 or 2) versus those with <1 year of seizure freedom (SSS > 2).

### 2.3.1 | Gene variant annotation and analysis

VCF files were annotated within the SAIL Databank using ANNOVAR (version: 2019Oct24)[23] with gene-based and filter-based annotations.[24] Genetic variants were defined as rare if the allele frequency was <.001 in the gnomAD exome collection (v2.1.1) and they did not occur more than twice within the study group. We

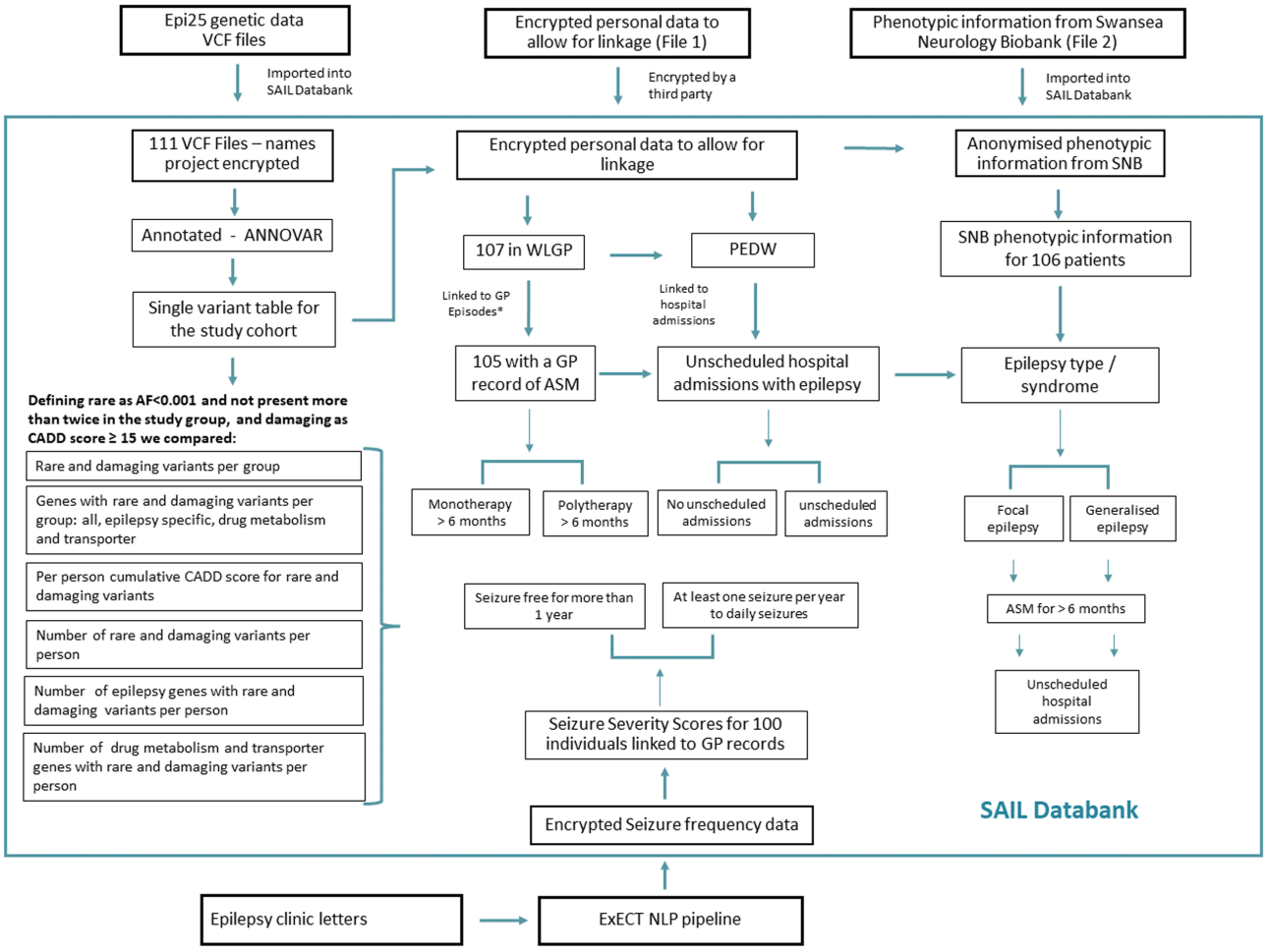


**FIGURE 1** Genetic data linkage and analysis within the Secure Anonymised Information Linkage (SAIL) Databank. The green rectangle represents data within the SAIL Databank. *Ninety-nine had a diagnosis of epilepsy in GP records, but all individuals in the cohort had epilepsy, as their diagnosis was confirmed before the Epi25 submission. AF, allele frequency; ANNOVAR, variant annotation tool; ASM, antiseizure medication; CADD, Combined Annotation-Dependent Depletion; ExECT, Extraction of Epilepsy Clinical Text; GP, general practitioner; NLP, natural language processing; PEDW, Patient Episode Database for Wales; SNB, Swansea Neurology Biobank; WLGP, Welsh Longitudinal General Practice dataset.

defined genetic variants as damaging if their Combined Annotation-Dependent Depletion (CADD) score was ≥15.[25]

### *Exome-wide burden analysis*

We defined the exome-wide gene burden score as the cumulative sum of CADD PHRED scores for all rare variants for that individual.[8] We also counted the number of rare and damaging variants for each individual.[26]

### *Gene-based analysis*

For each individual, we noted the number of rare and damaging variants in genes (1) associated with epilepsy, that is, epilepsy genes, neurodevelopment-associated epilepsy genes, and epilepsy-related genes[5,8,27] (Data S1); and (2) associated with drug metabolism and/or drug transporters, that is genes encoding phase 1 drug-metabolizing enzymes and genes encoding drug transporters[28] (Data S1). We compared Residual Variation Intolerance Scores (RVISs) in each of the groups identifying the most intolerant genes.[29]

We performed the exome-wide and gene-based analysis for the three outcome groups.

We used R Statistical Software (v4.1.2) for statistical analysis and graphs within the SAIL Databank.[30] We used the Wilcoxon rank sum test when comparing the groups.

## 3 | RESULTS

We linked 107 SNB patient exomes with SAIL health care data. A total of 105 had a primary care ASM prescription record, 104 had epilepsy classification information

(Table 1), and all were of White Western European ethnicity. The earliest GP event year relating to epilepsy (diagnosis or ASM prescription) was 2000, and the latest was 2019, giving a study window of 19 years (mean follow-up per patient = 12 years, range = 2–19 years).

## 3.1 | Hospital admissions

There was no significant difference in the exome-wide or gene-based analysis in individuals with no unscheduled epilepsy hospital admissions ($n = 79$) compared to those with unscheduled epilepsy admissions ($n = 28$; Figure 2A). A total of 562 (7.1 per person) and 168 (6.0 per person) unique qualifying (rare and damaging) variants in 501 and 145 genes were identified in the no admissions and the admissions groups, respectively. Variants in two genes associated with epilepsy, *CACNA1C* and *KCNQ1*, were exclusively present in the admissions group (both affecting fewer than five individuals). Qualifying variants were seen in three drug metabolism and transporter genes, with *CYP2D6* variants being present exclusively in the no admissions group (Figure 3). *CACNA1C* and *KCNQ1* are among the top 3% of genes intolerant to damaging variants (RVIS ranking), with *CYP2D6* being among the most tolerant 96%.

## 3.2 | ASM therapy

There was no significant difference, in the exome-wide or gene-based analysis, between people who were on ASM monotherapy ($n = 32$) and polytherapy ($n = 73$; Figure 2B). A total of 249 (7.8 per person) and 469 (6.4 per person) unique qualifying variants in 214 and 415 genes were identified in the monotherapy and polytherapy groups, respectively. There were a number of variants for epilepsy-associated genes exclusive to both groups, with the polytherapy group having more genes from among the top 3% of genes intolerant to damaging variants (RVIS ranking). There were no specific drug metabolism and transporter genes with qualifying variants that were uniquely associated with polytherapy (Figure 3). Two intolerant genes *CACNA1C* and *KCNQ1* (in the top 3% of intolerant genes) were present in all three groups, the unscheduled admissions, polytherapy, and not seizure-free groups.

## 3.3 | Seizure frequency

SSSs were calculated for 100 study participants (see Materials and Methods). Of those, 10 individuals were seizure-free for >1 year (SSS = 1 or 2), and 90 were experiencing seizures (at least one seizure per year to daily seizures; SSS > 2). There was no significant difference between the groups in the exome-wide analysis. Sixty-two (6.2 per person) and 684 (7.6 per person) unique qualifying variants in 62 and 586 genes were identified in the seizure-free and not seizure-free groups, respectively (Figure 3). One epilepsy-associated gene was found in the seizure-free group and 15 in the non seizure-free group, including CACNA1C, KCNQ1. Two less tolerant genes (top 4%), SCN5A and TRAK1, are also present in the polytherapy group. Variants of *CHD2* and *KCNH2* were present exclusively in the not seizure-free group, both from the top 4% of the least tolerant genes. A small number of variants of *ABCG2* were present exclusively in the not seizure-free group.

**TABLE 1** Total study and outcome group characteristics.

| | | Admissions | | ASM therapy[b] | | Seizure frequency[c] | |
|---|---|---|---|---|---|---|---|
| Characteristic | All | Admissions | No admissions | Polytherapy | Monotherapy | >1 year seizure freedom | <1 year seizure freedom |
| Total | 107 | 28 (26%) | 79 (74%) | 73 (69%) | 32 (31%) | 10 (10%) | 87 (90%) |
| Mean age, years | 41 | 40 | 41 | 42 | 37 | 41 | 41 |
| Male | 47 (44%) | 9 (32%) | 38 (48%) | 32 (44%) | 14 (44%) | 4 (40%) | 36 (41%) |
| Female | 60 (56%) | 19 (68%) | 41 (52%) | 41 (56%) | 18 (56%) | 6 (60%) | 51 (59%) |
| Focal epilepsy[a] | 73 (70%) | 19 (68%) | 54 (71%) | 54 (76%) | 18 (58%) | 5 (56%) | 61 (72%) |
| Generalized epilepsy[a] | 31 (30%) | 9 (32%) | 22 (29%) | 17 (24%) | 13 (42%) | 4 (44%) | 24 (28%) |

*Note*: Age was calculated for the end of the study.

Abbreviations: ASM, antiseizure medication; SAIL, Secure Anonymised Information Linkage; SNB, Swansea Neurology Biobank.

[a]There was no record of ASM prescriptions for two individuals, so there are 105 individuals in the ASM therapy outcomes group.

[b]Seizure frequency scores were extracted for 100 individuals (see Materials and Methods), but only 97 could be linked within the SAIL Databank to sex/age data, and 94 had diagnostic information in SNB (66 focal and 28 generalized).

[c]Epilepsy classification was not available for three individuals.

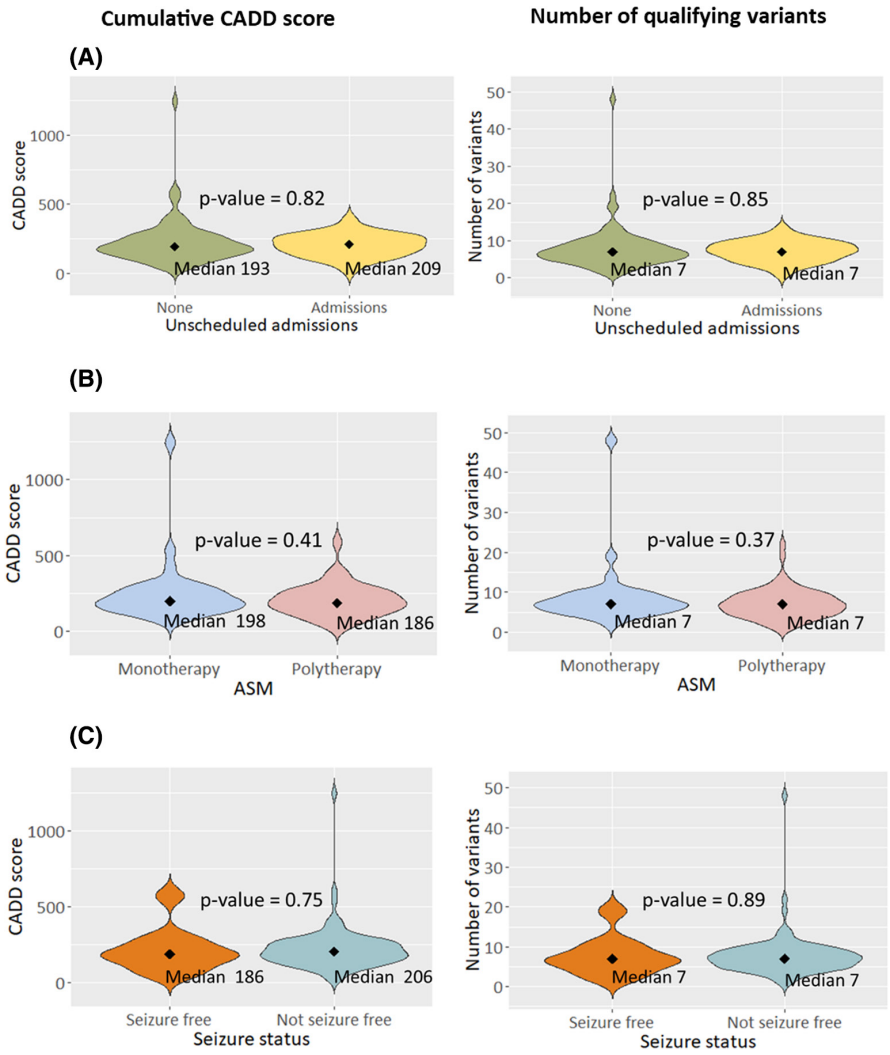


**FIGURE 2** Violin plots for cumulative Combined Annotation-Dependent Depletion (CADD) score for qualifying variants (left) and the number of qualifying variants (right) for (A) people with epilepsy and no unscheduled hospital admissions (green) and people with epilepsy and hospital admissions (yellow), (B) people on antiseizure medication (ASM) monotherapy for epilepsy (light blue) compared to those on polytherapy (light purple), and (C) individuals who are seizure-free (orange) compared to those who are not seizure-free (blue–gray). Qualifying variants are defined as rare and damaging (Materials and Methods). The width of the plots represents the probability density, and medians are shown on the graphs.

# 4 | DISCUSSION

## 4.1 | Main findings

We were able to import, annotate, and link genetic data, NLP-derived seizure frequency, and routinely collected electronic health care records for people with epilepsy within the SAIL Databank trusted research environment. This is a novel pathfinder project that has overcome several significant methodological hurdles and capability refinements. The pipeline developed in this project has the potential for other epilepsy (and other disease) studies. Using unscheduled hospital admissions, ASM polytherapy, and active seizures as a measure of more difficult to control epilepsy, we investigated the genetic burden of individuals with epilepsy by direct linkage with their routinely collected health care data.

We did not find any significant differences, in terms of overall or gene-specific burden of rare and damaging genetic variants, between our three epilepsy outcome groups. We found two genes with qualifying variants present in the unscheduled admissions, polytherapy, and active seizures groups that are intolerant to damaging variants, namely *CACNA1C* and *KCNQ1*.

*CACNA1C* encodes the alpha-1 subunit of a voltage-dependent L-type calcium channel expressed in human heart and brain.[31,32] Pathogenic variants in *CACNA1C* have been associated with a variety of phenotypes, including cardiac rhythm disorders as well as neurodevelopmental disorders including epilepsy and epileptic encephalopathies.[31,32] Voltage-gated calcium channels are targets for ASMs, and a *CACNA1C* haplotype has been previously associated with drug-resistant epilepsy.[33] *KCNQ1* encodes a voltage-gated potassium channel, predominately expressed in cardiac tissue but also expressed in the brain. Pathogenic *KCNQ1* variants can cause both long QT syndrome and epilepsy.[34–36] *ABCG2* is an ATP-binding cassette transporter primarily associated with breast cancer[37] that has also been investigated in the context of drug-resistant epilepsy, with no association being found.[38,39] A systematic review and meta-analysis did not identify an association between *ABCG2* polymorphisms

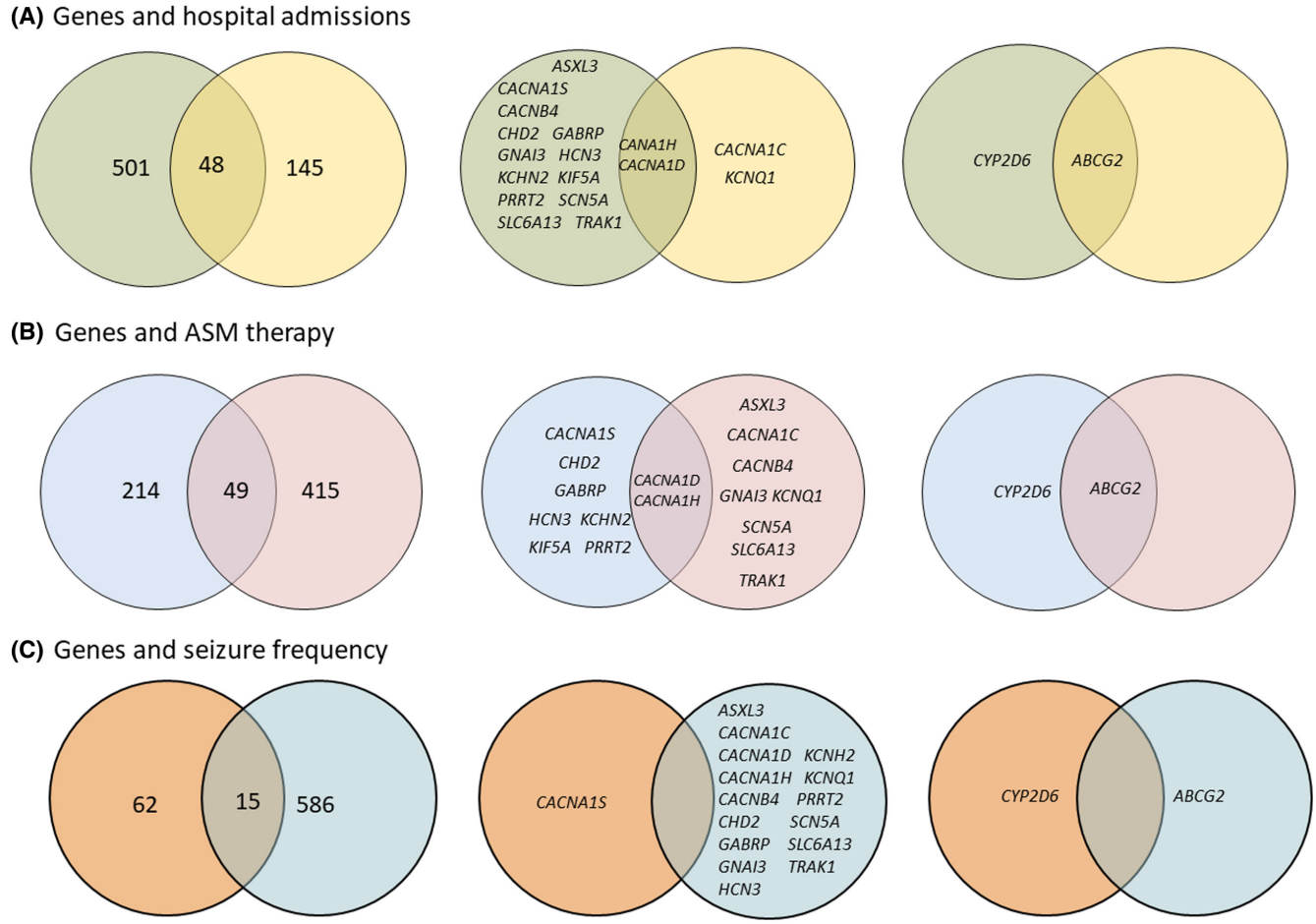


**FIGURE 3** Genes with qualifying (rare and damaging) variants in (A) those without unscheduled hospital admissions (green) compared to those with (yellow), (B) those on monotherapy (light blue) compared to those on polytherapy (light purple), and (C) those who are seizure-free (orange) compared to those who are not seizure-free (blue–gray). The Venn diagrams show the genes with qualifying variants in each outcome group as well as those common to both groups for all genes (first column), epilepsy genes (second column), and drug metabolism and transporter genes (third column).

and ASM response,[40] but a recent case–control study of drug-resistant epilepsy in children suggested an association with a specific variant.[41] There have not been reports of an association between the variants found in our cohort of not seizure-free individuals and epilepsy.

## 4.2 | Comparison to other studies

A study from the Electronic Medical Records and Genomics initiative used an algorithm to identify children with drug-resistant epilepsy (excluding structural causes and syndromic epilepsies) from routinely collected data.[42] They compared results from a drug transporter gene panel (82 genes) from 96 children with drug-resistant epilepsy, 343 children with drug-responsive epilepsy, and 896 controls. They found an association between *NTRK2* variants and drug-resistant epilepsy.[42] There were no qualifying *NTRK2* variants in our analysis. To our knowledge, no other studies have linked genetic data to epilepsy outcomes derived from routinely collected health records.

Studies investigating genetic burden and epilepsy outcomes have used clearly defined and differing seizure improvement measures to differentiate between ASM responders and nonresponders.[43,44] These studies are therefore difficult to compare to our results, but apart from findings relating to individual variants and specific ASMs,[45] no polygenic risk for drug resistance relating to seizure response has been identified.[46–48]

## 4.3 | Study strengths

We used well-phenotyped data with epilepsy specialist-confirmed diagnoses and subtypes. Seizure frequency data are not normally available in electronic health records, and we were able to use an NLP pipeline to add

seizure frequency information from clinic letters. We have created a gene variant annotation, filtering, and data linkage pipeline within the SAIL Databank trusted research environment that paves the way for future upscaled studies. The facility to reannotate variant files within the trusted research environment allows for reannotation without further data uploads when new variant datasets become available. Using routinely collected data allows for participants' information to be updated over time without any additional costs and with lower risk of loss to follow-up. Our work enables future population-level genetic studies using routinely collected data in both epilepsy and other diseases. Lastly, the quality, accuracy, and coverage of the gene variants are well provisioned through the Broad Institute Epi25 collaborative and are methodologically robust to exclude false variant calls.

### 4.4 | Study weaknesses

Our study was based on a small sample of individuals and is underpowered for gene variant association analysis. The sample size was limited by the number of participants with appropriate consent for genetic data linkage within the SAIL Databank. The characteristics of the patient group are not representative of the general epilepsy population (e.g., a greater proportion of people with drug-resistant epilepsy and no children). Using WES rather than whole-genome sequencing (WGS) is less powerful in discovering disease variants and misses non-coding expression, splicing, and epigenetic variation.[49]

In our study, stratification by epilepsy type or syndrome was not possible due to cohort size. We appreciate, however, that epilepsy syndromes and/or epilepsy types are themselves associated with epilepsy outcomes as well as having differing genetic variant profiles.[3,5] Untangling this association between genetic variation associated with epilepsy type/syndrome and outcomes will require a significantly larger, well-phenotyped epilepsy cohort.

At the same time, it is important to remember that many people with epilepsy do not have a specific epilepsy syndrome or subtype and understanding the broader effects of genetic variation is important in clinical practice.

Our outcome measures are only an approximation for more difficult to control epilepsy; some people may be seizure-free on polytherapy, whereas some people may have multiple seizures on monotherapy. Hospital admissions may reflect other issues, such as access to care/services, not just seizure/epilepsy severity, and we have not accounted for comorbidities in our analysis. Our measures of seizure frequency rely on documentation within clinic letters, and the numbers of clinic letters available for each individual in the cohort were not the same. It might be expected that people with more frequent seizures are reviewed more frequently in a specialist clinic, resulting in more detailed seizure frequency records.[50]

## 5 | CONCLUSIONS

We successfully uploaded, annotated, and linked genetic sequence data to anonymized health care records, augmenting the data with seizure frequency information from NLP. We established a pathway that can be followed by other studies, using different health indicators, or linking to other data sources. Our study did not detect a genetic influence on real-world epilepsy outcomes and requires a larger WES/WGS sample size.

### AUTHOR CONTRIBUTIONS

Mark I. Rees, William O. Pickrell, Seo-Kyung Chung, Kerina H. Jones, and Beata Fonferko-Shadrach contributed to the concept and study design. Beata Fonferko-Shadrach, William O. Pickrell, Seo-Kyung Chung, and Mark I. Rees drafted the manuscript and figures. Beata Fonferko-Shadrach, Huw Strafford, Carys Jones, and Arron S. Lacey contributed to data acquisition and analysis. Ashley Akbari, Ronan A. Lyons, David Ford, Simon Thompson, Mark Baker, and Robert Powell contributed to data acquisition. All authors approved the final draft version.

### ACKNOWLEDGMENTS

This work was supported by staff funded as part of the Wales Gene Park, an infrastructure support group funded by the Welsh Government through Health and Care Research Wales. The views expressed are those of the authors and not necessarily those of Health and Care Research Wales or the Welsh Government. We thank and acknowledge our colleagues in the Broad Institute and the Epi25 collaborative (http://epi-25.org/). Epi25 regards this work as a local secondary study limited to repatriated WES data. No other submitting Epi25 sites were included due to the limited study access to Wales National Health Service health care data. This study used anonymized data held in the Secure Anonymised Information Linkage Databank. We would like to acknowledge all the data providers who make anonymized data available for research. Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australian University Librarians.

## CONFLICT OF INTEREST STATEMENT

None of the authors has any conflict of interest to disclose.

## ORCID

*Beata Fonferko-Shadrach* https://orcid.org/0000-0003-3282-9502
*Huw Strafford* https://orcid.org/0000-0003-2147-3969
*Ashley Akbari* https://orcid.org/0000-0003-0814-0801

## REFERENCES

1. World Health Organization. WHO | epilepsy: a public health imperative. 2019.
2. Thijs RD, Surges R, O'Brien TJ, Sander JW. Epilepsy in adults. Lancet. 2019;393(10172):689–701. https://doi.org/10.1016/S0140-6736(18)32596-0
3. Ellis CA, Petrovski S, Berkovic SF. Epilepsy genetics: clinical impacts and biological insights. Lancet Neurol. 2020;19:93–100. https://doi.org/10.1016/S1474-4422(19)30269-8
4. The International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun. 2018;9(1):5269. https://doi.org/10.1038/s41467-018-07524-z
5. Epi25 Collaborative. Electronic address: s.berkovic@unimelb.edu.au; Epi25 Collaborative. Ultra-rare genetic variation in the epilepsies: a whole-exome sequencing study of 17,606 individuals. Am J Hum Genet. 2019;105(2):267–82. https://doi.org/10.1016/j.ajhg.2019.05.020
6. Naimo GD, Guarnaccia M, Sprovieri T, Ungaro C, Conforti FL, Andò S, et al. A systems biology approach for personalized medicine in refractory epilepsy. Int J Mol Sci. 2019;20(15):1–15. https://doi.org/10.3390/ijms20153717
7. Margari L, Legrottaglie AR, Vincenti A, Coppola G, Operto FF, Buttiglione M, et al. Association between SCN1A gene polymorphisms and drug resistant epilepsy in pediatric patients. Seizure. 2018;55:30–5. https://doi.org/10.1016/J.SEIZURE.2018.01.002
8. Leu C, Balestrini S, Maher B, Hernández-Hernández L, Gormley P, Hämäläinen E, et al. Genome-wide polygenic burden of rare deleterious variants in sudden unexpected death in epilepsy. EBioMedicine. 2015;2(9):1063–70. https://doi.org/10.1016/j.ebiom.2015.07.005
9. Kansu B, Pickrell WO, Lacey AS, Edwards F, Samolia G, Rees MI, et al. Long-term outcomes after epilepsy surgery, a retrospective cohort study linking patient-reported outcomes and routine healthcare data. Epilepsy Behav. 2020;111:107196. https://doi.org/10.1016/j.yebeh.2020.107196
10. Lee-Lane E, Torabi F, Lacey A, Fonferko-Shadrach B, Harris D, Akbari A, et al. Epilepsy, antiepileptic drugs, and the risk of major cardiovascular events. Epilepsia. 2021;62(7):1604–16. https://doi.org/10.1111/epi.16930
11. Myland M, Buysse B, Tsong W, Power GS, Nordli D, Chin R. Seizure frequency, healthcare resource utilisation and mortality in childhood epilepsy: a retrospective cohort study using the THIN database. Arch Dis Child. 2019;104(11):1070–6. https://doi.org/10.1136/archdischild-2019-316910
12. Symonds JD, McTague A. Epilepsy and developmental disorders: next generation sequencing in the clinic. Eur J Paediatr Neurol. 2020;24:15–23. https://doi.org/10.1016/j.ejpn.2019.12.008
13. Striano P, Minassian BA. From genetic testing to precision medicine in epilepsy. Neurotherapeutics. 2020;17(2):609–15. https://doi.org/10.1007/s13311-020-00835-4
14. Epi25 Collaborative. 2014 [cited 2022 May 1] http://epi-25.org/genes-in-epilepsy.
15. Jones KH, Ford DV, Thompson S, Lyons RA. A profile of the SAIL databank on the UK secure research platform. Int J Popul Data Sci. 2019;4(2):1134. https://doi.org/10.23889/IJPDS.V4I2.1134
16. Lyons RA, Jones KH, John G, Brooks CJ, Verplancke JP, Ford DV, et al. The SAIL databank: linking multiple health and social care datasets. BMC Med Inform Decis Mak. 2009;9(1):1–8. https://doi.org/10.1186/1472-6947-9-3
17. Benson T. The history of the read codes: the inaugural James read memorial lecture 2011. Inform Prim Care. 2012;19(3):173–82. https://doi.org/10.14236/jhi.v19i3.811
18. Ford DV, Jones KH, Verplancke JP, Lyons RA, John G, Brown G, et al. The SAIL databank: building a national architecture for e-health research and evaluation. BMC Health Serv Res. 2009;9:157. https://doi.org/10.1186/1472-6963-9-157
19. Fonferko-Shadrach B, Lacey AS, Roberts A, Akbari A, Thompson S, Ford DV, et al. Using natural language processing to extract structured epilepsy data from unstructured clinic letters: development and validation of the ExECT (extraction of epilepsy clinical text) system. BMJ Open. 2019;9(4):e023232. https://doi.org/10.1136/bmjopen-2018-023232
20. Fonferko-Shadrach B, Lacey AS, Strafford H, Jones KH, Jones C, Baker M, Chung SK, Pickrell WO, Rees MI. Investigating the relationship between anti-seizure medications and seizure frequency recorded in epilepsy clinic letters using natural language processing; poster presentation. ILAE British branch annual scientific conference, abstract book, Cardiff. London, UK: ILAE British Branch; 2022. p. 96.
21. Fitzgerald MP, Kaufman MC, Massey SL, Fridinger S, Prelack M, Ellis C, et al. Assessing seizure burden in pediatric epilepsy using an electronic medical record-based tool through a common data element approach. Epilepsia. 2021;62:1617–28. https://doi.org/10.1111/epi.16934
22. Fonferko-Shadrach B, Lacey AS, White CP, Powell HWR, Sawhney IMS, Lyons RA, et al. Validating epilepsy diagnoses in routinely collected data. Seizure. 2017;52:195–8. https://doi.org/10.1016/j.seizure.2017.10.008
23. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164. https://doi.org/10.1093/nar/gkq603
24. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc. 2015;10(10):1556–66. https://doi.org/10.1038/nprot.2015.105
25. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47(D1):D886–94. https://doi.org/10.1093/nar/gky1016
26. Wang G-S, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet. 2007;8(10):749–61. https://doi.org/10.1038/nrg2164
27. Wang J, Lin ZJ, Liu L, Xu HQ, Shi YW, Yi YH, et al. Epilepsy-associated genes. Seizure. 2017;44:11–20. https://doi.org/10.1016/j.seizure.2016.11.030

28. Ahmed S, Zhou Z, Zhou J, Chen SQ. Pharmacogenomics of drug metabolizing enzymes and transporters: relevance to precision medicine. Genomics Proteomics Bioinformatics. 2016;14(5):298–313. https://doi.org/10.1016/j.gpb.2016.03.008
29. Petrovski S, Gussow AB, Wang Q, Halvorsen M, Han Y, Weir WH, et al. The intolerance of regulatory sequence to genetic variation predicts gene dosage sensitivity. PLoS Genet. 2015;11(9):e1005492. https://doi.org/10.1371/JOURNAL.PGEN.1005492
30. R Core Team. R: a language and environment for statistical computing. 2021.
31. Bozarth X, Dines JN, Cong Q, Mirzaa GM, Foss K, Lawrence Merritt J II, et al. Expanding clinical phenotype in CACNA1C related disorders: from neonatal onset severe epileptic encephalopathy to late-onset epilepsy. Am J Med Genet Part A. 2018;176(12):2733–9. https://doi.org/10.1002/ajmg.a.40657
32. Rodan LH, Spillmann RC, Kurata HT, Lamothe SM, Maghera J, Jamra RA, et al. Phenotypic expansion of CACNA1C-associated disorders to include isolated neurological manifestations. Genet Med. 2021;23(10):1922–32. https://doi.org/10.1038/s41436-021-01232-8
33. Lv N, Qu J, Long H, Zhou L, Cao Y, Long L, et al. Association study between polymorphisms in the CACNA1A, CACNA1C, and CACNA1H genes and drug-resistant epilepsy in the Chinese Han population. Seizure. 2015;30:64–9. https://doi.org/10.1016/j.seizure.2015.05.013
34. MacCormick JM, McAlister H, Crawford J, et al. Misdiagnosis of Long QT syndrome as epilepsy at first presentation. Ann Emerg Med. 2009;54(1):26–32. https://doi.org/10.1016/j.annemergmed.2009.01.031
35. Goldman AM, Glasscock E, Yoo J, Chen TT, Klassen TL, Noebels JL. Arrhythmia in heart and brain: KCNQ1 mutations link epilepsy and sudden unexplained. Sci Transl Med. 2009;1(2):2ra6. https://doi.org/10.1126/scitranslmed.3000289
36. Tiron C, Campuzano O, Pérez-Serra A, Mademont I, Coll M, Allegue C, et al. Further evidence of the association between LQT syndrome and epilepsy in a family with KCNQ1 pathogenic variant. Seizure. 2015;25:65–7. https://doi.org/10.1016/j.seizure.2015.01.003
37. Kowal J, Ni D, Jackson SM, Manolaridis I, Stahlberg H, Locher KP. Structural basis of drug recognition by the multidrug transporter ABCG2. J Mol Biol. 2021;433(13):1–48. https://doi.org/10.1016/j.jmb.2021.166980
38. Kim DW, Lee SK, Chu K, Jang IJ, Yu KS, Cho JY, et al. Lack of association between ABCB1, ABCG2, and ABCC2 genetic polymorphisms and multidrug resistance in partial epilepsy. Epilepsy Res. 2009;84(1):86–90.
39. Kwan P, Wong V, Ng PW, Lui CHT, Sin NC, Wong KS, et al. Gene-wide tagging study of the association between ABCC2, ABCC5 and ABCG2 genetic polymorphisms and multidrug resistance in epilepsy. Pharmacogenomics. 2011;12(3):319–25.
40. Zan X, Yue G, Hao Y, Sima X. A systematic review and meta-analysis of the association of ABCC2/ABCG2 polymorphisms with antiepileptic drug responses in epileptic patients. Epilepsy Res. 2021;175:106678. https://doi.org/10.1016/j.eplepsyres.2021.106678
41. Mousavi SF, Hasanpour K, Nazarzadeh M, Adli A, Bazghandi MS, Asadi A, et al. ABCG2, SCN1A and CYP3A5 genes polymorphism and drug-resistant epilepsy in children: a case-control study. Seizure. 2022;97:58–62. https://doi.org/10.1016/j.seizure.2022.03.009
42. Almoguera B, McGinnis E, Abrams D, Vazquez L, Cederquist A, Sleiman PM, et al. Drug-resistant epilepsy classified by a phenotyping algorithm associates with NTRK2. Acta Neurol Scand. 2019;140(3):169–76. https://doi.org/10.1111/ane.13115
43. Heavin SB, McCormack M, Wolking S, Slattery L, Walley N, Avbersek A, et al. Genomic and clinical predictors of lacosamide response in refractory epilepsies. Epilepsia Open. 2019;4(4):563–71. https://doi.org/10.1002/epi4.12360
44. Wolking S, Schulz H, Nies AT, McCormack M, Schaeffeler E, Auce P, et al. Pharmacoresponse in genetic generalized epilepsy: a genome-wide association study. Pharmacogenomics. 2020;21(5):325–35. https://doi.org/10.2217/pgs-2019-0179
45. Glauser TA, Holland K, O'Brien VP, Keddache M, Martin LJ, Clark PO, et al. Pharmacogenetics of antiepileptic drug efficacy in childhood absence epilepsy. Ann Neurol. 2017;81(3):444–53. https://doi.org/10.1002/ana.24886
46. Wolking S, Moreau C, Mccormack M, Krause R. Assessing the role of rare genetic variants in drug-resistant, non-lesional focal epilepsy. Ann Clin Transl Neurol. 2021;8:1376–87. https://doi.org/10.1002/acn3.51374
47. Wolking S, Moreau C, Nies AT, Schaeffeler E, McCormack M, Auce P, et al. Testing association of rare genetic variants with resistance to three common antiseizure medications. Epilepsia. 2020;61(4):657–66. https://doi.org/10.1111/epi.16467
48. Perucca P, Stanley K, Harris N, McIntosh AM, Asadi-Pooya AA, Mikati MA, et al. Rare genetic variation and outcome of surgery for mesial temporal lobe epilepsy. Ann Neurol. 2023;93(4):752–61. https://doi.org/10.1002/ANA.26581
49. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci U S A. 2015;112(17):5473–8. https://doi.org/10.1073/pnas.1418631112
50. Baker M, Fonferko-Shadrach B, Stafford H, Bain S, Howell OW, Sawhney IM, Powell RH, Rees MI, Pickrell WO. An open-access biobank for epilepsy research and patient impact; poster presentation. ILAE British branch scientific conference, abstract book, Cardiff. London, UK: ILAE British Branch; 2022. p. 56.


## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Fonferko-Shadrach B, Lacey AS, Strafford H, Jones C, Baker M, Powell R, et al. Genetic influences on epilepsy outcomes: A whole-exome sequencing and health care records data linkage study. Epilepsia. 2023;00:1–10. https://doi.org/10.1111/epi.17766