



RESEARCH ARTICLE

Development of an Innovative SQL-Based Approach to Identify Potential Patients with Neurotransmitter Disorders

Emily Fox¹, Vishal Mehta², Rajesh Madhu³, Evangeline Wassmer⁴, Ruchi Arora⁵, Tony Cox^{6*}, Dave Heaton⁶, Julia Granerod⁶ and Mark Rance¹

¹Medical Affairs, PTC Therapeutics, Guildford, UK

²Paediatric Neurology, Hull University Teaching Hospitals NHS Trust, Hull, UK

³Neurosciences, Alder Hey Children's NHS Foundation Trust, Liverpool, UK

⁴Neurology, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK

⁵Paediatric Neurology, Norfolk and Norwich University Hospital NHS Foundation Trust, Norwich, UK

⁶Real-World Evidence, OPEN Health, Marlow, UK

*Corresponding author: Tony Cox, Real-World Evidence, OPEN Health, Marlow, UK



Abstract

Background: The neurotransmitter disorders (NTDs) signs and symptoms range from early-onset severe neurological manifestations, often accompanied with developmental delay, to later-onset moderate movement disorders [1]. Timely diagnosis remains challenging.

Methods: To develop the SQL algorithm we analysed pseudonymised patient records from the Hospital Episode Statistics (HES) database, which covers all National Health Service inpatient admissions in England, between April 2010 and May 2020. Data extracted included diagnoses, clinical procedures, clinical specialty, and tariff codes. Records with NTD diagnoses (International Statistical Classification of Diseases and Related Health Problems 10th Revision codes E708-E709) were extracted. Clinical codes were ordered by frequency and, using expert clinical input, scored between one and five based on the degree of association with NTDs. To help characterise potentially undiagnosed patients, those records with a total score ≥ 53 were flagged as potential NTDs, and the five most common codes were then used as further selection criteria.

Results: A total of 55 million patients were admitted to English hospitals in the 10-year study period; 1,972 records had a weighted score of ≥ 53 . The most common code recorded in these records was paediatric neurology, followed by epilepsy, dystonia, cerebral palsy, and hypotonia. A total of 1,469 patients, which formed the case review population, had a score ≥ 53 , attended paediatric neurology, were diagnosed with epilepsy, and had a diagnosis of dystonia, cerebral palsy, or hypotonia.

Conclusions: An innovative SQL coding approach using de-identified hospital records profiled potential patients with NTDs for further testing. This derived from the hospital patient administration system (PAS) has implications for shortening the time to definitive diagnosis and enabling earlier intervention. Our methodology may be applied to other rare diseases or under diagnosed conditions.

Keywords

Neurotransmitter disorder, Aromatic L-amino acid decarboxylase, structured (lower case) query language, diagnosis

What is known about the subject?

- The incidence of neurotransmitter disorders (NTDs) in the general population remains unknown.
- Timely diagnosis of patients remains challenging due to nonspecific symptoms and lack of available metabolite measurements required to diagnose these disorders.
- An accurate diagnosis is important as some NTDs are readily treatable.

What this study adds

- An innovative SQL algorithm was developed to



Citation: Fox E, Mehta V, Madhu R, Wassmer E, Arora R, et al. (2022) Development of an Innovative SQL-Based Approach to Identify Potential Patients with Neurotransmitter Disorders. Int J Rare Dis Disord 5:048. doi.org/10.23937/2643-4571/1710048

Accepted: September 14, 2022; **Published:** September 16, 2022

Copyright: © 2022 Fox E, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

query hospital Patient Administration Systems (PAS) to profile potential patients with NTDs and flag them for further testing by clinicians. The algorithm was devised following analysis of HES data and testing with expert clinical input.

- This has the potential to shorten the time to definitive diagnosis, enabling earlier intervention.
- The methodology used may apply to the detection of other rare diseases or under diagnosed conditions

Introduction

Neurotransmitter disorders (NTDs) result from genetic defects of neurotransmitter metabolism and transport and include defects of catecholamine, serotonin, bipterin, glycine, pyridoxine, and gamma amino butyric acid metabolism [2]. Signs and symptoms often manifest in early childhood and include hypotonia, movement disorders, autonomous dysregulations, and impaired development [2]. The incidence of NTDs in the general population remains unknown. A retrospective cohort study carried out at a single centre in Canada reported a 4% prevalence of inherited NTDs in patients with global developmental delay, neonatal hypotonia, neonatal seizures, epilepsy, and movement disorders, who underwent cerebrospinal fluid (CSF) neurotransmitter measurements for diagnostic work-up [2]. Measurement of CSF neurotransmitter metabolites (e.g., homovanillic acid [HVA], 5-hydroxyindolacetic acid [5-HIAA], 3-O-methyldopa [3-OMD], tetrahydrobiopterin, bipterin, and neopterin) is used to identify disorders of catecholamine, serotonin, and bipterin metabolisms, with diagnosis confirmed by direct sequencing of candidate genes [3].

Aromatic L-amino acid decarboxylase deficiency (AADC-d), an NTD characterised by impaired synthesis of the catecholamines dopamine, adrenaline and noradrenaline, and serotonin, has been described in over 135 patients thus far the medical literature [4]. A founder mutation has been identified in East Asia with the prevalence of AADC-d highest in Taiwan (1:32,000); however, the disease is not specific to those of Asian ethnicity and also occurs in individuals of other descent [5,6]. Most patients present in early infancy with a severe phenotype including early onset hypotonia, oculogyric crises, ptosis, dystonia, hypokinesia, impaired development, and autonomic dysfunction; however, mild disease can occur [7]. Low levels of 5-HIAA and HVA, and high concentrations of 3-OMD, L-Dopa, and 5-OH tryptophan (5-HTP) are often seen in the CSF of patients with AADC-d [7,5]. AADC-d can be genetically confirmed in the vast majority of patients [7].

Timely diagnosis of patients with NTDs remains challenging as the nonspecific constellation of symptoms may overlap with those of other neurological syndromes, for example cerebral palsy and epileptic

encephalopathy's [1]. Under-diagnosis may also result from lack of lumbar punctures (LPs) and lack of available metabolite measurements required to diagnose these disorders [2]. An accurate diagnosis is important as some NTDs are readily treatable [8]. Misdiagnosis or delayed diagnosis may result in an irreversible decline in the patient's condition and incur significant costs to the healthcare system [9]. In the Canadian study, 5.3% of patients who underwent LP for the measurement of CSF HVA and 5-HIAA had one of the treatable inherited metabolic disorders with favourable short-term neurodevelopment outcomes highlighting the importance of an early and specific diagnosis [1]. Supplementation of lacking neurotransmitter precursors or restoration of deficient cofactors for endogenous enzymatic synthesis forms the basis of treatment for NTDs [7]. Treatment strategies for AADC-d often vary between centres and treatment response in patients is often disappointing; however, a small proportion of patients with AADC-d respond well to L-Dopa or a dopamine agonist [2].

This study aimed to develop structured language query (SQL)-based approach using electronic health record data from PAS to identify potential patients with NTDs/AADC-d across hospitals in England who should receive proper diagnostic workup.

Methods

Data source

The cohort was derived from Hospital Episode Statistics (HES), pseudonymised electronic medical records from all National Health Service (NHS) inpatient and outpatient admissions in England [10]. Each HES episode has up to 20 diagnoses, recorded using International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) codes [11]. Procedures are recorded using OPCS Classification of Interventions and Procedures, and Healthcare Resource Group (HRG) codes to group patient events that consume a similar level of resource. These data were utilised to build the SQL select query for clinicians to utilise at local level via a secure on-line portal. These data were only available for clinicians and clinical review. Ethical and research approvals were not required for this study as it is a permitted re-use of Hospital Episode Statistics and patient identifiable information was not utilized.

Study population

Records from between April 2010 and May 2020 were retrieved from HES. Data included diagnoses, clinical procedures, clinical specialty, and tariff codes. Data from the following patient cohorts were extracted and used to develop the SQL SELECT query:

E708-709 cohort: This included records from patients with an ICD-10 code for E708 or E709 admitted during

the 10-year study period. There are no specific codes for NTDs/AADC-d within the ICD-10, and patients with NTDs/AADC-d are likely captured under either the E708 ('Other disorders of aromatic amino-acid metabolism') or E709 ('Disorder of aromatic amino-acid metabolism, unspecified') codes.

Total HES cohort: This included records from all patients admitted to all hospitals in England during the study period.

Identification of trigger points

Two key papers were identified through a targeted literature review and by clinical experts and screened for content on important clinical features associated with NTDs/AADC-d [7,12]. An initial list of key features was extracted and subsequently reviewed by two pediatric neurologists and two pediatricians to ensure all relevant signs/symptoms, procedures, and clinical specialties associated with an NTD patient were captured. The resulting list of codes (n = 115) constituted a trigger point table comprising the most relevant events a potential NTD patient might undergo prior to diagnosis (see Table 1 in online supplement for full list). The final list of agreed trigger points was translated into HES terminology (ICD-10, OPCS, and HRG codes) for use in subsequent analyses.

Development of the SQL query

The E708-709 cohort represented the dataset used to devise a SQL SELECT query. The pre-defined trigger

points were applied to this cohort to identify patients with at least one of the 115 codes present in their record. The combinations of the occurrences of the trigger points were identified and likely NTD/ AADC-d patients were identified within the wider four-digit ICD-10 codes. The combination of codes present within these patient records were further analyzed by clinical experts to assess the likelihood of an NTD and it was agreed that these represented the potential patient profile. The 115 trigger point codes were subsequently ranked and given a weighted score between one and five based on the degree of likely association with NTDs (i.e., a score of five indicated high association with NTDs; Table 1). A total score was calculated for each patient record containing at least one trigger point by adding together the weighted score for each trigger point code that appeared in the patient record. A score of ≥ 53 was deemed an appropriate cut off for patients highly likely to have NTDs/AADC-d.

Patient and public involvement

Patients, their families and members of the public were not involved in the design of the study or in setting the parameters of the SQL query.

Application of SQL SELECT query

HES data testing the query

The SQL SELECT query was applied to the entire HES database (i.e., total HES cohort) and all patients with a score of ≥ 53 were analyzed. All codes present within

Table 1: NHS Trusts with ≥ 20 potential patients with NTDs.

Trust code	Trust name	Total patients
RBS	Alder Hey Children's NHS Foundation Trust	50
R1H	Barts Health NHS Trust	40
RVV	East Kent Hospitals University NHS Foundation Trust	40
RJ1	Guy's and St Thomas' NHS Foundation Trust	40
RDE	East Suffolk and North Essex NHS Foundation Trust	35
RVR	Epsom and St Helier University Hospitals NHS Trust	30
RC9	Bedfordshire Hospitals NHS Foundation Trust	25
RQ3	Birmingham Women's and Children's NHS Foundation Trust	25
RXR	East Lancashire Hospitals NHS Trust	25
RJ2	Lewisham and Greenwich NHS Trust	25
RAJ	Mid and South Essex NHS Foundation Trust	25
REF	Royal Cornwall Hospitals NHS Trust	25
RAL	Royal Free London NHS Foundation Trust	25
RCU	Sheffield Children's NHS Foundation Trust	25
RGT	Cambridge University Hospitals NHS Foundation Trust	20
RQM	Chelsea and Westminster Hospital NHS Foundation Trust	20
RJZ	King's College Hospital NHS Foundation Trust	20
RGN	North West Anglia NHS Foundation Trust	20
RDR	Sussex Community NHS Foundation Trust	20
RA7	University Hospitals Bristol and Weston NHS Foundation Trust	20
RTG	University Hospitals of Derby and Burton NHS Foundation Trust	20

the records of these patients were aggregated and ordered by frequency, and the five most common key trigger points were listed. Patients most likely to have NTDs/AADC-d, or the case review population, included those with a total score of ≥ 53 and combinations of the key trigger points. The outputs of this analysis led to a key table highlighting the differing combinations of codes and specifically the number of expected patients at each hospital trust site in England.

Implementation in a real world setting

The SQL SELECT query based on the HES analysis was modified for use on the NHS Trust Patient Administration System (PAS). This enables a simple searching capability across the central trust databases, is safer for the Trusts, and supports which in turn is governance. The modified SQL SELECT query was run internally within the Trust to maintain patient confidentiality. Patient identifiable information was not received by OPEN Health or PTC Therapeutics at any stage. The list of identified patients was sent directly to the relevant clinician, and it was left to their discretion to decide which patients might benefit from further diagnostic workup.

Following detailed discussion with the Medicines and Healthcare products Regulatory Agency (MHRA) in November 2021, the MHRA decided that the SQL SELECT query is a general purpose product which is customized by the end user to select patients in the HES database. Therefore, this would not be a medical device as the SQL SELECT query does not provide decision making capability. All decisions regarding diagnosis and management for specific patients remain with their responsible Healthcare Professional.

Results

HES data

A total of 55 million patients were admitted to English hospitals between April 2010 and May 2020, of which 20,350,842 had at least one of the 115 trigger points listed in their record. The total weighted score in these patients ranged from one to 110. Almost three quarters ($n = 14,717,481$, 72.3%) had a total weighted score ≤ 3 , and 1,972 patients had a weighted score of ≥ 53 and were noted as possible NTDs/AADC-d patients. The most common code recorded in patients with a score ≥ 53 was for attendance to pediatric neurology, followed by diagnostic codes for epilepsy, dystonia, cerebral palsy, and hypotonia.

Of the 1,972 patients with a score ≥ 53 , 1822 attended pediatric neurology and were diagnosed with epilepsy. A further 1,469 had a diagnosis of dystonia, cerebral palsy, or hypotonia; these formed the case review population and were deemed highly likely to have NTDs/AADC-d. Twenty-one NHS Trusts were identified with ≥ 20 potential NTD patients (Table 1).

Two hundred and eighty-nine patients in HES had

an E708-709 diagnostic code during the study period. The 1,469 patients highly likely to have NTDs/AADC-d indicated by the SQL SELECT query was 80% more than captured by the E708-709 codes alone.

Trust data

So far, the modified SQL SELECT query has been successfully run in twelve NHS Trusts and health boards and has identified up to 130 patients suitable for further clinical review.

Seven potential NTDs/AADC-d patients were identified in one Trust for which data were obtained. Upon case review, one patient had symptoms consistent with an NTD, however, the 3-OMD measurement used as an initial screen for AADC-d had not been carried out. Thus, this one patient ($n = 1/7$, 14%) might have AADC-d and would benefit from further testing as per consensus guidelines [7].

Discussion

We developed an innovative programming language approach to indicate patients across England with potential NTDs/AADC-d that would benefit from further diagnostic workup. This SQL SELECT query could provide clinicians the opportunity to improve patient outcomes through earlier NTD detection and subsequent intervention.

We noted 1,469 patients with potential NTDs/AADC-d across all NHS hospitals in England, 80% more than those with an E708-709 code. Our SQL SELECT query was based on a profile of signs, symptoms, and treatment specialty which are more straight forward to code than an NTD diagnosis. Rare diseases such as NTDs which present with non-specific symptoms and are a challenge to diagnose might be miscoded (for example as a mimicker syndrome) in HES. Upon further review of seven patients identified within one Trust as potential NTDs/AADC-d, one (14%) did not have an alternative diagnosis, had not undergone complete investigation, and thus needs further testing. This confirms the SQL SELECT query is working and has significant implications for patient management and outcome.

Methodology involving a SQL-based approach has not previously been applied to the diagnosis of NTDs. The slightly different approach of machine learning however, has been used successfully for the diagnosis of other rare diseases [13]. Machine learning algorithms build models based on sample data in order to make predictions or decisions. It should be noted that this is not what was conducted in this study. Rather, a SQL-based approach was used to manage data incorporating relations among entities and variables and retrieve relevant data.

Application of a SQL-based approach may provide advantages over other methods of disease detection, especially with regards to rare diseases. Due to their

infrequent occurrence, rare diseases may not initially be suspected by treating physicians. In addition, rare diseases often have non-specific presentations and genetic components which may require specialized testing, thus further complicating diagnosis. Another advantage is that the SQL query was developed on the whole HES database which represents the entire population of England, rather than a selected database from a single centre or a few centers in one area which may introduce biases. HES includes patients of all ages and although symptom onset in AADC-d occurs early in life, diagnosis may not occur until years later even after the age of 20 years [5]. Thus, restricting the data by age may result in some cases being missed. A further advantage of this SQL-based approach is the potential to select all patients with a rare disease such as NTDs with important implications for powering further studies, including those assessing outcome.

Some limitations however need consideration. SQL-based methodology can be subject to biases, including misclassification, errors in measurement, and missing data. Thus, clinical judgment should be used alongside the application of these techniques. Miscoding could have occurred in HES; however, data quality has been shown to be better in later years [14]. Further work includes the further identification of potential patients in Trust data and diagnostic workup of these patients to validate the SQL query.

In conclusion, the approach described in this study highlights probable cases of NTDs/AADC-d for further diagnostic workup. This SQL-based methodology may improve detection of NTDs/AADC-d in clinical settings, allowing for earlier intervention and improved patient outcomes. Our methodology may be applied to other rare diseases and under diagnosed conditions and serve as a novel strategy to aid patient identification.

Funding

This work was fully funded by PTC Therapeutics, Ltd for the purpose of medical education and improving patient care. No healthcare professional was paid by PTC Therapeutics, Ltd as part of this project.

Grant/Award Number

Not applicable.

Competing Interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: No support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Author Contributions

The study concept was devised by PTC Therapeutics, Ltd. Records were retrieved by Harvey Walsh as part of the OPEN Health Group. The data were analyzed by Harvey Walsh. The manuscript was written by Dr Julia Granerod. All authors read, commented, and approved the manuscript.

Acknowledgments

These studies are funded by PTC Therapeutics, Ltd.

References

- Brennenstuhl H, Jung-Klawitter S, Assmann B, Opladen T (2019) Inherited disorders of neurotransmitters: Classification and practical approaches for diagnosis and treatment. *Neuropediatrics* 50: 2-14.
- Mercimek-Mahmutoglu S, Sidky S, Hyland K, Patel J, Donner EJ, et al. (2015) Prevalence of inherited neurotransmitter disorders in patients with movement disorders and epilepsy: A retrospective cohort study. *Orphanet J Rare Dis* 10: 12.
- Cordeiro D, Bullivant G, Cohn RD, Raiman J, Mercimek-Andrews S (2018) Outcome of patients with inherited neurotransmitter disorders. *Can J Neurol Sci* 45: 571-576.
- Pearson TS, Gilbert L, Opladen T, Garcia-Cazorla A, Mastrangelo M, et al. (2020) AADC deficiency from infancy to adulthood: Symptoms and developmental outcome in an international cohort of 63 patients. *J Inherit Metab Dis* 43: 1121-1130.
- Hyland K, Reott M (2020) Prevalence of aromatic l-amino acid decarboxylase deficiency in at-risk populations. *Pediatr Neurol* 106: 38-42.
- Chien YH, Chen PW, Lee NC, Hsieh WS, Chiu PC, et al. (2016) 3-O-methyl dopa levels in newborns: Result of newborn screening for aromatic l-amino-acid decarboxylase deficiency. *Mol Genet Metab* 118: 259-263.
- Wassenberg T, Molero-Luis M, Jeltsch K, Hoffmann GF, Assmann B, et al. (2017) Consensus guideline for the diagnosis and treatment of aromatic l-amino acid decarboxylase (AADC) deficiency. *Orphanet J Rare Dis* 12.
- Siu WK (2015) Genetics of monoamine neurotransmitter disorders. *Transl Pediatr* 4: 175-180.
- Opladen T, Cortès-Saladefont E, Mastrangelo M, Horvath G, Pons R, et al. (2016) The International Working Group on Neurotransmitter related Disorders (iNTD): A worldwide research project focused on primary and secondary neurotransmitter disorders. *Mol Genet Metab Rep* 9: 61-66.
- Hospital episode statistics (HES). NHS Digital.
- (2010) International statistical classification of diseases and related health problems 10th revision. ICD-10 Version: 2010.
- Himmelreich N, Montioli R, Bertoldi M, Carducci C, Leuzzi V, et al. (2019) Aromatic amino acid decarboxylase deficiency: Molecular and metabolic basis and therapeutic outlook. *Mol Genet Metab* 127: 12-22.
- Schaefer J, Lehne M, Schepers J, Prasser F, Thun S (2020) The use of machine learning in rare diseases: A scoping review. *Orphanet J Rare Dis* 15.
- Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, et al. (2018) Understanding hospital episode statistics (HES). CLOSER Resource Report.