Molecular Genetics and Metabolism Reports 40 (2024) 101101

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism Reports

journal homepage: www.elsevier.com/locate/ymgmr

Estimating prevalence of classical homocystinuria in the United States using Optum's de-identified market clarity data

Mahim Jain [a,*], Mehul Shah [b], Kamlesh M. Thakker [c], Andrew Rava [d], Agness Pelts Block [b], Colette Ndiba-Markey [d], Lionel Pinto [b,1]

[a] Kennedy Krieger Institute, Johns Hopkins Medicine, Baltimore, MD, USA
[b] Travere Therapeutics, Inc., San Diego, CA, USA
[c] Notting Hill Consulting LLC, Celebration, FL, USA
[d] Genesis Research Group, Hoboken, NJ, USA

ARTICLE INFO

Keywords:
Classical homocystinuria
Prevalence
Total homocysteine

ABSTRACT

Background and objectives: Prevalence estimates for classical homocystinuria (HCU) are variable and likely underestimated due to underdiagnosis. Claims data represent a strong but seldom used resource to analyze prevalence of HCU. The aim of this study was to estimate a prevalence range of HCU in the US utilizing a combination of diagnosis codes, total homocysteine levels, and clinical presentations indicative of HCU.

Methods: This was a non-interventional retrospective cohort study, using Optum's de-identified Market Clarity Data, with a patient identification period from January 01, 2016, through September 30, 2021. An algorithm was developed to identify 2 cohorts of patients using broad and strict definitions of HCU. The index date was the date within the identification period on which the first criterion was met for the inclusion criteria. Baseline demographics, clinical characteristics, and complications were assessed and summarized using descriptive statistics. Crude and standardized prevalence estimates were calculated.

Results: There were 3880 and 633 patients that met the relevant inclusion criteria for the broad and strict cohorts, respectively. The projected US prevalence of HCU was calculated to be 17,631 and 3466 based on the broad and strict definitions, respectively. The average annual standardized prevalence across 2016–2020 was 5.29 and 1.04 per 100,000 people for the broad and strict cohorts, respectively.

Conclusions: Prevalence estimates of HCU vary depending on databases or datasets used and identification criteria. Many patients with clinical presentations suggesting a diagnosis of HCU did not have an associated diagnosis, potentially indicating underdiagnosis or underreporting. Future research should study alternative methods, such as the identification algorithm in our analysis, to better diagnose and understand the true prevalence of HCU.

1. Introduction

Classical homocystinuria (HCU), also known as cystathionine beta-synthase (CBS)-deficient homocystinuria, is an autosomal recessive genetic inborn error of metabolism [1–3]. It is due to a deficiency in the CBS enzyme, resulting in elevated levels of homocysteine and methionine (Met) [1,3]. A clinical diagnosis of HCU is made on the basis of clinical features, biomarkers including total homocysteine (tHcy), and additional confirmation via genetic testing. Normal tHcy levels are typically below 15 μM, while tHcy levels above 100 μM are usually suggestive of HCU [3–6]. A 2020 study found that tHcy levels above 30 μM may also be indicative of HCU in some patients, especially those with undiagnosed HCU [7].

HCU is associated with risks of complications, including thrombotic/thromboembolic events, cognitive impairment and developmental delays, ectopia lentis and myopia, and skeletal changes, including osteopenia, osteoporosis, pectus excavatum, and elongated arms and legs (marfanoid habitus) [1–3]. Early treatment and control of tHcy levels can prevent or limit complications [8].

The worldwide prevalence has been reported to be from 1:200,000 to

* Corresponding author at: Division of Orthogenetics Nemours Children's Hospital Wilmington, DE 19803, USA.
E-mail address: mahim.jain@nemours.org (M. Jain).
[1] At the time of the study.

https://doi.org/10.1016/j.ymgmr.2024.101101
Received 31 January 2024; Received in revised form 7 March 2024; Accepted 28 May 2024
Available online 7 June 2024

1:335,000 [3]. However, there are also geographical differences, with the highest estimated incidence in Qatar at 1:1800 [6,9], due to a founder mutation. Prior US prevalence estimates were based predominantly on newborn screening and reported to be about 1 per 100,000-200,000 [1,10], but other methods of detection have resulted in different prevalence estimates [7,11].

The prevalence of HCU is variable and likely underestimated due to underdiagnosis as a result of multiple factors [12]. Newborn screening (NBS) for HCU began in some US states in 1968, but was not recommended in all US states until 2009 [13]. While screening for HCU now occurs in all US NBS programs [14], the method used is not sensitive and results in a high percentage of missed diagnoses [3,15,16]. First-tier NBS assesses for elevated Met, however patients with the milder form may not have elevated Met levels at the time NBS is performed [3,11,12,14,17]. Additionally, many symptoms of HCU are not specific to the disease (eg, cognitive impairment and skeletal abnormalities) and further contribute to delayed or missed diagnoses.

The true prevalence of HCU in the US remains unknown and many patients may be undiagnosed. Limited published research using administrative claims data is available on identifying patients with HCU beyond the diagnosis code, and claims data represent a strong but seldom used resource to analyze the prevalence of HCU. To get a better estimate of the prevalence of HCU, there is a need to apply a more robust strategy that combines various modalities (eg, claims data and clinical parameters such as tHcy levels, clinical features, and complications) to identify the overall universe of likely patients with HCU.

The aim of this study was to estimate a prevalence range of HCU in the US using Optum's de-identified Market Clarity Data and by utilizing multiple potential means of identifying HCU, including diagnosis codes, homocysteine levels, and clinical presentations.

2. Methods

This was a non-interventional US retrospective cohort study with a patient identification period from January 01, 2016, through September 30, 2021.

2.1. Data source

Optum's de-identified Market Clarity Data (Market Clarity) was used for determination of the concepts related to homocystinuria. The Market Clarity dataset links electronic health record data with historical, linked administrative claims data, pharmacy claims, physician claims, and facility claims (with clinical information) and is inclusive of medications prescribed and administered. Clinically rich and specific data elements sourced from the electronic health record include lab results, vital signs and measurements, diagnoses, procedures, and information derived from unstructured clinical notes using natural language processing. Natural Language Processed (NLP) Concepts are identified and created based on broad topics such as Medications; Signs, Disease, and Symptoms (SDS); Measurements; Observations; etc. The data are harvested from the notes fields within the Electronic Medical Records provided to Optum from over 50 large health care systems throughout the US.

The Market Clarity dataset is fully Health Insurance Portability and Accountability Act (HIPAA)-compliant and contains de-identified data. Institutional review board approval was not required for this study.

2.2. Study population and participants

While an International Classification of Disease, Tenth Revision (ICD-10) homocystinuria-related code (E72.11) exists, there is not an ICD-10 code specific for HCU. Furthermore, within the Market Clarity dataset, we found that among patients with an E72.11 ICD-10 code, there were limited data on tHcy testing and values (Supplementary Fig. 1). This suggested that simply looking at ICD codes was not sufficient to effectively classify individuals with HCU. Therefore, an algorithm was developed to identify 2 cohorts of patients using broad and strict definitions of HCU (Fig. 1). Patient cohorts were identified in a stepwise method, based on the presence of a homocystinuria-related ICD-10 code (E72.11) or a homocystinuria SDS NLP term, followed by the patients' highest tHcy level at any time during the study period. For patients with mildly elevated tHcy, potential causes of elevated tHcy other than HCU (such as disorders of cobalamin metabolism and methylenetetrahydrofolate reductase [MTHFR] deficiency) [4,6,18] were also considered when creating the patient identification algorithm (Fig. 1). Potential secondary causes of elevated tHcy used as exclusion criteria, including a list of ICD codes used, are included in the Supplementary Material. These exclusion criteria included diagnoses such as chronic kidney disease, diabetes, hypothyroidism, nutritional anemia, vitamin B12 deficiency anemia, other megaloblastic anemia, and MTHFR deficiency. In addition to clinical characteristics, phenotypic expressions were used to further refine the cohort selection. A phenotypic expression was considered to be present if one or more of the associated diagnosis codes were present. Full details on the patient identification algorithm are described in the Supplementary Material. The index date was the date within the identification period on which the first criterion was met for the inclusion criteria. Baseline patient characteristics were assessed in the 6 months prior to index. This does not apply to the prevalence calculations.

2.3. Variables and outcomes

Baseline demographics assessed included index year (2016 to 2021), age at index (in years), age group at index (<18, 18–44, 45–64, 65–74, and ≥ 75 years), gender (male, female, unknown or missing), region (Northeast, Midwest, South, West, Other/Unknown), Hispanic origin (yes, no, unknown), race (African American, Asian, White, other/unknown), and health insurance type (commercial, Medicare, Medicaid, other payer type, uninsured, unknown).

Clinical characteristics assessed included highest tHcy level and highest Met level (at any time during the study period). Baseline complications/comorbidities assessed included the Charlson comorbidity index (CCI), Charlson comorbidities, thrombotic/thromboembolic events including deep vein thrombosis (DVT), pulmonary embolism (PE), portal vein thrombosis, renal vein thrombosis, myocardial infarction (MI), and cerebrovascular thromboembolic events (stroke/transient ischemic attack [TIA]), ocular events (retinal detachment, lens dislocation, myopia), and skeletal events (osteoporosis, scoliosis, pectus excavatum, pectus carinatum, pes cavus, genu valgum, marfanoid habitus).

Annual and across-the-study (2016–2020) estimates of the crude and standardized prevalence of HCU per 100,000 US population were reported for both the broad and strict cohorts.

2.4. Statistical analysis

Patient demographics, clinical characteristics, baseline complications, and comorbidities were summarized using descriptive statistics. Categorical variables were summarized using frequencies and percentages. Continuous variables were summarized using means (standard deviations [SDs]) and percentiles (minimum, 1st quartile [Q1], median, 3rd quartile [Q3], and maximum). Missing data were considered a separate category in the prevalence analysis and were described using frequency counts and percentages for both categorical and continuous covariates. Crude and standardized prevalence estimates with 95% confidence intervals (CIs) were reported. Crude prevalence was calculated as the total number of prevalent patients (patients with a HCU diagnosis as defined by our identification algorithm) during the year of interest divided by the total number of patients in the dataset who contributed ≥1 person-day during the year of interest. Standardized prevalence was estimated using data from the US Census Bureau 2020 and directly standardized using the age, gender, and race/ethnicity

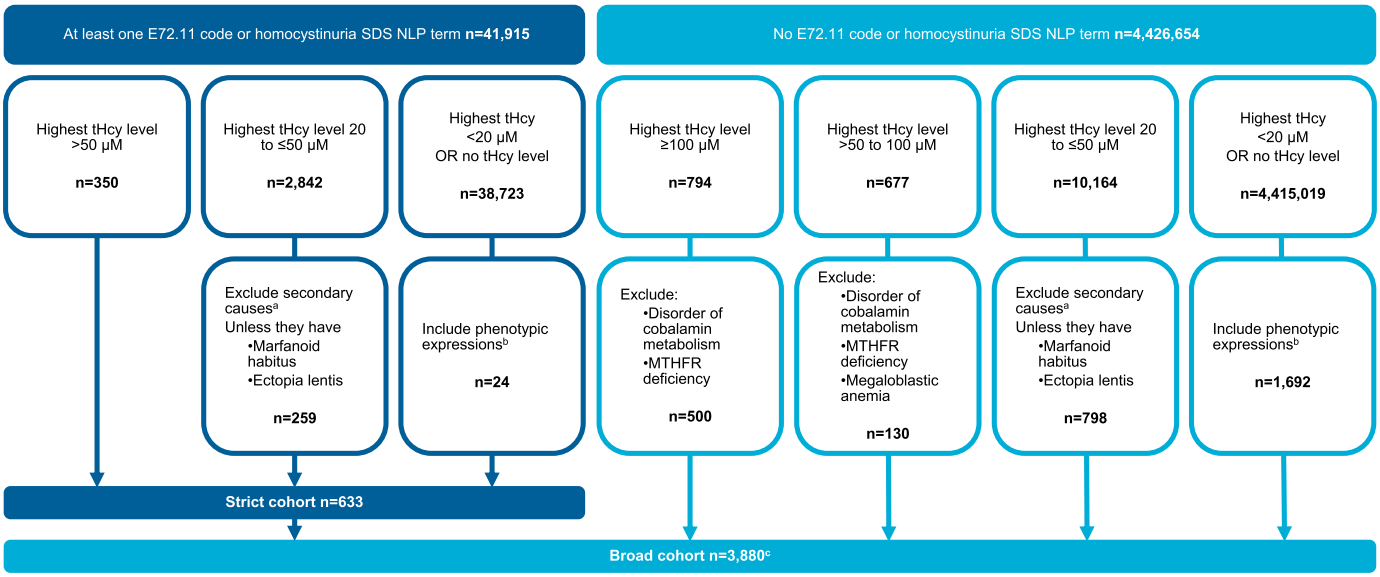


Fig. 1. Classical Homocystinuria Patient Identification Algorithm.
Abbreviations: CKD, chronic kidney disease; ESKD, end-stage kidney disease; mo, month; MTHFR, methylenetetrahydrofolate reductase; NLP, Natural Language Processed; SDS, signs, disease, and symptoms; tHcy, total homocysteine.
[a]**Secondary causes; At any time:** Megaloblastic anemia, disorder of cobalamin metabolism, folate deficiency, CKD, ESKD, renal transplant, diabetes, hypothyroidism; **Within 12 mo:** Myocardial infarction.
[b]**Phenotypic expressions:** 1. Ectopia lentis AND (cerebrovascular thrombotic/thromboembolic event OR neurologic feature) **exclude**: Marfanoid habitus, sulfite oxidase deficiency (E72.19); 2. Pectus excavatum AND (cerebrovascular thrombotic/thromboembolic event OR [any thrombotic/thromboembolic event AND neurologic feature]) **exclude**: Marfanoid habitus, sulfite oxidase deficiency (E72.19); 3. Marfanoid habitus AND cerebrovascular thrombotic/thromboembolic event AND neurologic feature AND (ectopia lentis OR pectus excavatum) **exclude**: Sulfite oxidase deficiency (E72.19).
[c]The total number of patients in the broad cohort includes 3753 patients, as shown above, plus 127 patients with or without E72.11 or SDS NLP term and 2 records of betaine at least 12 months apart (excluding disorder of cobalamin metabolism and MTHFR deficiency).

distributions. Average annual estimates were calculated as the average of the standardized estimates from 2016, 2017, 2018, 2019, and 2020. Annual prevalence was not calculated for 2021 because we did not have a full year of data in the dataset. The total US population estimate from the US Census Bureau 2022 was used to extrapolate the standardized prevalence estimates to the current US population.

3. Results

3.1. Demographics and baseline characteristics

Based on the study criteria applied to the dataset, a total of 3880 patients met the relevant inclusion criteria for the broad cohort and 633 patients met the relevant inclusion criteria for the strict cohort (Fig. 1). Baseline demographics for both cohorts are shown in Table 1.

In the broad cohort (n = 3880), the mean age was 57.2 years and there was a nearly even gender distribution (48.0% female). Most patients were 18 years and older at index (95.1%). The largest number of patients in the broad cohort were between 45 and 64 years of age (33.9%). The majority of patients were White (82.9%), 9.4% were African American, and 1.0% were Asian. Close to one-half (41.4%) of the patients in the broad cohort had commercial insurance, while 36.9% and 16.6% were covered by Medicare or Medicaid, respectively (Table 1).

In the strict cohort (n = 633), the mean age was 50.0 years and there was a nearly even gender distribution (46.6% female). Most patients were 18 years and older at index (94.8%). The largest number of patients in the strict cohort were between 45 and 64 years of age (42.2%). The majority of patients were White (79.3%), 12.6% were African American, and 1.4% were Asian. Approximately one-half (49.8%) of the patients in the strict cohort had commercial insurance. Most of the remainder of patients had either Medicaid or Medicare (23.1% and 22.9%, respectively) (Table 1).

3.2. Clinical characteristics

Maximum homocysteine levels are outlined in Table 2. The mean (SD) maximum tHcy level for the broad cohort was 50.1 (57.0) µM. The median (Q1, Q3) maximum tHcy level was 27.3 (21.8, 60.0) µM. Of the patients with a tHcy level in the broad cohort (n = 1700, 43.8% of the total 3880 in the broad cohort), the majority had a maximum tHcy level between 20 µM to <50 µM (62.2%) followed by 31.9% with a level of 50 µM or greater. The mean (SD) maximum Met level for the broad cohort was 109.2 (247.0) µM.

The mean (SD) CCI for the broad cohort was 1.2 (1.9) (Table 3). Cerebrovascular and peripheral vascular diseases were the most common HCU-related baseline Charlson comorbidities reported in the broad cohort (n = 3880) (18.1% and 13.2%, respectively) (Table 3). In addition, thrombotic/thromboembolic events during the baseline period were reported in 21.7% of patients in the broad cohort (who had at least 1 acute event), followed by 9.8% for skeletal conditions and 3.9% for ocular disorders (Table 3 and Supplementary Table 1).

The mean (SD) maximum tHcy level for the strict cohort was 69.8 (77.9) µM and the median (Q1, Q3) maximum tHcy level was 52.5 (24.8, 81.5) µM (Table 2). Of the patients with a tHcy level in the strict cohort (n = 582, 91.9% of the total 633 in the strict cohort), the majority (54.8%) had a maximum tHcy level of 50 µM or greater, followed by 44.3% with a level between 20 µM to <50 µM (Table 2). The mean (SD) maximum Met level for the strict cohort was 339.4 (429.4) µM. The mean (SD) CCI for the strict cohort was 0.7 (1.4) (Table 3). Cerebrovascular and peripheral vascular diseases were the most common HCU-related Charlson comorbidities reported in the strict cohort (n = 633) (both at 10.9%) (Table 3). In addition, thrombotic/thromboembolic events during the baseline period were reported in 21.5% of patients in the strict cohort (who had at least 1 acute event), followed by 6.3% for skeletal conditions and 3.3% for ocular disorders (Table 3 and Supplementary Table 1).

Table 1
Baseline Demographics of the Broad and Strict Classical Homocystinuria Cohorts[a].

	Broad cohort	Strict cohort
Total, No. (%)	3880 (100.0)	633 (100.0)
Index year, No. (%)		
2016	1016 (26.2)	188 (29.7)
2017	899 (23.2)	156 (24.6)
2018	686 (17.7)	122 (19.3)
2019	583 (15.0)	84 (13.3)
2020	439 (11.3)	60 (9.5)
2021[b]	257 (6.6)	23 (3.6)
Age at index (continuous)		
Mean (SD)	57.2 (21.1)	50.0 (18.0)
Median (Q1, Q3)	60.0 (44.0, 74.0)	51.0 (39.0, 63.0)
Min-Max	0–89	0–87
Age (categorical), No. (%)		
<18	189 (4.9)	33 (5.2)
18–44	793 (20.4)	192 (30.3)
45–64	1315 (33.9)	267 (42.2)
65–74	642 (16.5)	92 (14.5)
≥75	941 (24.3)	49 (7.7)
Gender, No. (%)		
Female	1862 (48.0)	295 (46.6)
Male	2016 (52.0)	338 (53.4)
Unknown	2 (0.1)	0 (0.0)
Region, No. (%)		
Midwest	2116 (54.5)	344 (54.3)
Northeast	550 (14.2)	86 (13.6)
Other/Unknown	205 (5.3)	24 (3.8)
South	647 (16.7)	96 (15.2)
West	362 (9.3)	83 (13.1)
Hispanic origin, No. (%)		
Hispanic	150 (3.9)	32 (5.1)
Not Hispanic	3382 (87.2)	549 (86.7)
Unknown	348 (9.0)	52 (8.2)
Race, No. (%)		
African American	363 (9.4)	80 (12.6)
Asian	37 (1.0)	9 (1.4)
White	3216 (82.9)	502 (79.3)
Other/Unknown	264 (6.8)	42 (6.6)
Insurance type, No. (%)		
Commercial	1608 (41.4)	315 (49.8)
Medicaid	646 (16.6)	146 (23.1)
Medicare	1433 (36.9)	145 (22.9)
Other payer type	61 (1.6)	8 (1.3)
Uninsured	95 (2.4)	12 (1.9)
Unknown	37 (1.0)	7 (1.1)

Abbreviations: Q1, 1st quartile; Q3, 3rd quartile; SD, standard deviation.
[a] Because of rounding, percentages may not total 100.
[b] For 2021, results were included up to September.

Table 2
Maximum Homocysteine Levels in the Broad and Strict Classical Homocystinuria Cohorts[a,b,c].

	Broad cohort	Strict cohort
Total with at least 1 lab, No. (%)	1700 (43.8)	582 (91.9)
Total with ≥2 labs ≥50 μM, No. (%)	120 (7.1)	101 (17.4)
tHcy		
Mean (SD), μM	50.1 (57.0)	69.8 (77.9)
Median (Q1, Q3), μM	27.3 (21.8, 60.0)	52.5 (24.8, 81.5)
Min-Max, μM	2–877	4–877
Maximum tHcy, categorical, No. (%)		
<20 μM,	99 (5.8)	5 (0.9)
20 to <50 μM	1058 (62.2)	258 (44.3)
≥50 μM	543 (31.9)	319 (54.8)
50 to <100 μM	339 (19.9)	208 (35.7)
≥100 μM	204 (12.0)	111 (19.1)

Abbreviations: Q1, 1st quartile; Q3, 3rd quartile; SD, standard deviation; tHcy, total homocysteine.
[a] In patients with at least 1 lab value.
[b] At any time during the study period.
[c] Because of rounding percentages may not total 100.

Table 3
Baseline Comorbidities in the Broad and Strict Classical Homocystinuria Cohorts.

	Broad	Strict
Total, No.	3880	633
Charlson comorbidity index		
Mean (SD)	1.2 (1.9)	0.7 (1.4)
Median (Q1, Q3)	0.0 (0.0, 2.0)	0.0 (0.0, 1.0)
Min-Max	0–15	0–15
Charlson comorbidities, No. %		
Cerebrovascular disease	701 (18.1)	69 (10.9)
Peripheral vascular disease	514 (13.2)	69 (10.9)
Congestive heart failure	448 (11.5)	36 (5.7)
Myocardial infarction	201 (5.2)	18 (2.8)
Dementia	143 (3.7)	6 (0.9)
Hemiplegia or paraplegia	139 (3.6)	25 (3.9)
Occurrence of ≥ 1 acute event during the baseline period, No. (%)		
Baseline thrombotic/thromboembolic events[a]	841 (21.7)	136 (21.5)
Baseline skeletal conditions[b]	382 (9.8)	40 (6.3)
Baseline ocular disorder[c]	152 (3.9)	21 (3.3)

Abbreviations: Q1, 1st quartile; Q3, 3rd quartile; SD, standard deviation.
[a] Included deep vein thrombosis, pulmonary embolism, portal vein thrombosis, renal vein thrombosis, myocardial infarction, and cerebrovascular thromboembolic events (stroke/transient ischemic attack).
[b] Included osteoporosis, scoliosis, pectus excavatum, pectus carinatum, pes cavus, marfanoid habitus, and genu valgum.
[c] Included retinal detachment, lens dislocation, and myopia.

3.3. Patient identification

Among patients without a homocystinuria (E72.11) diagnosis code or SDS NLP term who had available tHcy levels ($n = 1496$), 42.1% had a highest value >50 μM and 33.4% had a highest value ≥100 μM. Of the 83.7% of patients who did not have a homocystinuria (E72.11) diagnosis code or SDS NLP term (3247 out of 3880), 52.1% had phenotypic expressions consistent with the manifestations of HCU (Fig. 1).

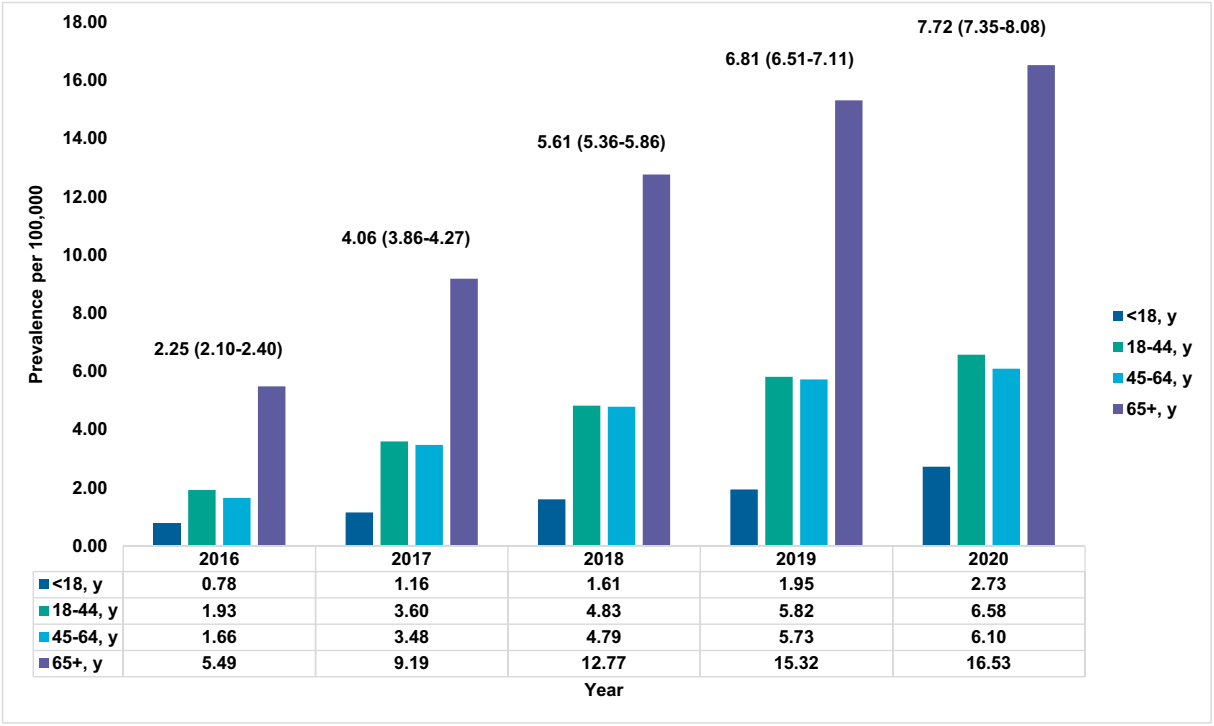
3.4. Prevalence

The projected prevalence of HCU in the US was 17,631 based on the broad definition and 3466 based on the strict definition. The average annual standardized prevalence across 2016–2020 was 5.29 per 100,000 people for the broad cohort and 1.04 per 100,000 people for the strict cohort. In the broad cohort, the annual prevalence ranged from 2.25 per 100,000 people in 2016 to 7.72 per 100,000 people in 2020 and in the strict cohort, the annual prevalence ranged from 0.46 per 100,000 people in 2016 to 1.53 per 100,000 people in 2020 (Fig. 2A and B).

4. Discussion

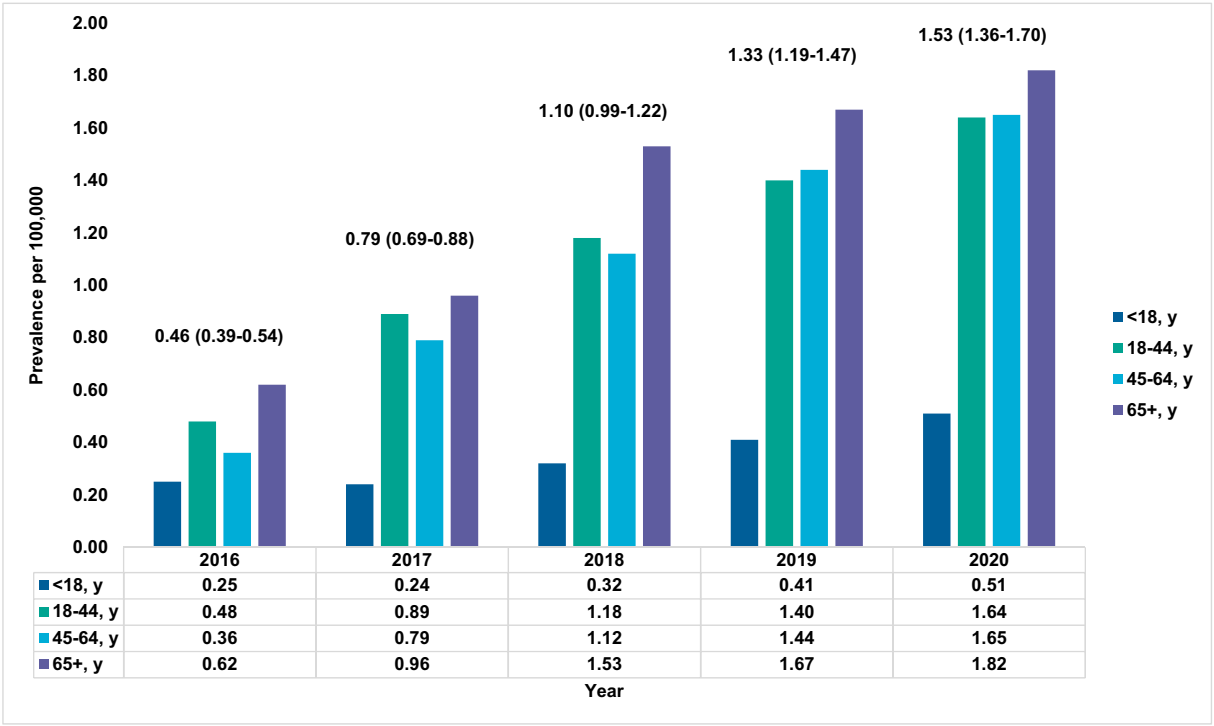
This study developed an algorithm to identify patients with HCU based on diagnosis codes, lab values, and clinical presentations using both broad and strict definitions of HCU. Prior similar studies have stratified patients only by ICD codes or tHcy levels [7,11]. Utilizing ICD codes alone would likely result in an over-estimate of the true prevalence. By excluding secondary causes and including associated clinical characteristics and phenotypic expressions, this algorithm therefore aimed to identify patients with HCU, even if a diagnosis for HCU was not recorded.

The demographics of the patients in our study were similar to what is known about patients with HCU. As expected, there were more patients meeting the broad definition of HCU compared with the strict definition. The largest number of patients in both the broad and strict cohorts were between the ages of 45 and 64, which is a similar finding to another study assessing the prevalence of patients with homocysteine levels above 30 μM [7], and another similar study assessing prevalence of



	2016	2017	2018	2019	2020
<18, y	0.78	1.16	1.61	1.95	2.73
18-44, y	1.93	3.60	4.83	5.82	6.58
45-64, y	1.66	3.48	4.79	5.73	6.10
65+, y	5.49	9.19	12.77	15.32	16.53

(a)



	2016	2017	2018	2019	2020
<18, y	0.25	0.24	0.32	0.41	0.51
18-44, y	0.48	0.89	1.18	1.40	1.64
45-64, y	0.36	0.79	1.12	1.44	1.65
65+, y	0.62	0.96	1.53	1.67	1.82

(b)

Fig. 2. Estimated Crude and Standardized Prevalence of Classical Homocystinuria per 100,000 People, by Age Group and Year for the Broad (A) and Strict (B) Cohorts.
Abbreviations: y, years. Crude prevalence estimates are shown in the table and bar graph. Standardized estimates with 95% CIs are shown above the bar graphs for each year.

patients with HCU using administrative claims data [11]. This similar database study comparing patients with HCU and phenylketonuria found a comparable age at index to our study, with strictly defined HCU patients a mean of 56.8 years old at index and those with broadly defined HCU a mean of 55.5 years old at index in their study [11]. Additionally, the demographics of both cohorts in our study were similar, except patients in the strict cohort were slightly younger. However, most patients in both cohorts were over 18 years old at index.

Due to the range in severity of symptoms and complications related to HCU, symptoms leading to a diagnosis do not always manifest in pediatric patients and can present in adults [3,6,16]. Additionally, the conditions assessed in our patient identification algorithm are not always phenotypically expressed in pediatric patients. Given this, along with the similar age demographics in our study and the aforementioned study using a different database, the disparity in prevalence by age could be related to the disease itself and not the specific databases used for the study. These results suggest that many patients with HCU are likely diagnosed at an older age or undiagnosed.

The strict cohort had higher mean maximum tHcy levels and more patients with maximum tHcy levels above 50 μM. These results are likely a reflection of identifying more patients with milder cases of HCU when using the broad definition and more patients with severe cases of HCU when using the strict definition. Similarly, only a small percentage of patients had Met lab results available but, of those, patients in the strict cohort had higher mean levels compared with those patients in the broad cohort, likely reflecting more severe HCU in the strict cohort. Many patients with high tHcy levels and clinical presentations indicative of HCU did not have a corresponding recorded diagnosis of HCU, suggesting possible underdiagnosis or underreporting.

Based on our study, using the July 2022 US Census Bureau estimate, the projected prevalence of HCU in the US is 17,631 based on the broad definition and 3466 based on the strict definition. Our estimates are slightly lower than those found in other similar studies which ranged from approximately 31,000 in a broadly defined cohort and 12,000 in the strictly defined cohort [11] or the estimated 31,000 patients with a tHcy level $>$ 30 μM [7]. These variations could be due to differences in the databases or datasets used for the studies, the study period, and differences in the definitions and identification of HCU. However, our prevalence estimates were validated in a similar study using the IQVIA PharMetrics Plus and Ambulatory Electronic Medical Record (AEMR) 2018–2021 databases from January 1, 2018, through May 31, 2022 [19]. The PharMetrics Plus database is comprised of fully adjudicated medical and pharmacy claims, while the AEMR comprises approximately 75 million US patient records from an opt-in provider research network [19]. In this study, a similar, but slightly lower projected prevalence of HCU in the US was found, 11,732 for the broad cohort and 2800 for the strict cohort [19]. The average annual standardized prevalence estimates (2018–2021) were also slightly lower, at 3.52 per 100,000 (broad cohort) and 0.84 per 100,000 (strict cohort) [19]. These estimates do need to be interpreted within the context of limitations associated with the IQVIA databases due to missing data and minimal overlap between the PharMetrics Plus and AEMR data. The results of these studies suggest that HCU prevalence estimates vary depending on factors such as the identification criteria, cohort definitions, and database or dataset used. In both our study and the aforementioned study [19], many patients with clinical presentations suggesting a diagnosis of HCU did not have an associated diagnosis of HCU, potentially indicating underdiagnosis or underreporting.

In our study, there was a trend noted of increasing prevalence over the time of the study. There are many uncertainties regarding HCU given that this is a rare disease. However, the cause of this is likely multifactorial. There could be increased awareness of HCU and its diagnosis. There was also a decrease in the number of patients available within the dataset over time, possibly due to the intervening COVID pandemic years resulting in fewer patients seeking medical care and more patients without insurance. Another potential contributing factor relates to the dataset used for the study. It is likely that more recent patients with better capture of claims are included in the dataset, leading to increased counts in recent years.

Future research should explore alternative methods to better diagnose HCU (such as the relatively robust algorithm detailed in our analysis) and understand its prevalence. The results of our study help to increase awareness for HCU and advocate for improved screening, recognition, and diagnosis of HCU. Identifying these patients earlier could potentially help improve morbidity, mortality, and cost of care for patients with HCU.

4.1. Limitations

This study was limited to data in the Market Clarity data, which included primarily commercially insured patients, and may not be representative of the broader US population. Standardization of the prevalence estimates to the US population was performed to minimize this bias. Missing data or errors in detection of homocystinuria-related ICD-10 codes and SDS NLP terms could introduce bias, including potential underestimation of the US prevalence of HCU. Additionally, patients with HCU may not be adequately captured in our dataset due to self-management through diet or vitamin intake.

There are unavoidable limitations related to claims-based datasets and databases and ascertainment bias is possible. This analysis was restricted to more recently available data (2016–2020), so for some patients, the index date may not indicate the date of first HCU diagnosis. As a result, patients could have been diagnosed at younger ages, prior to data availability for this study. Additionally, if patients had undiagnosed diseases or had diagnoses outside the database network, they would not be able to be ascertained. Given the lack of an existing ICD-10 code specific only to HCU, exclusion of other metabolic causes of elevated homocysteine with the same ICD-10 code (E72.11) could not be guaranteed. However, the patient identification algorithm used in this study aimed to accurately identify patients with HCU by using a combination of diagnosis codes, lab values, and clinical presentations.

The limited number of tHcy levels reported in the dataset used makes it difficult to assess if patients have consistently high tHcy levels. However, this could also be related to variations in clinician testing preferences, as they may only require one elevated tHcy test in combination with clinical symptoms consistent with HCU to make the diagnosis. Additionally, very few patients in both cohorts (<1%) had methionine levels available, which may be driving the average levels up.

The higher age of our study population is thought to be due to higher rates of underdiagnosis in the pediatric population. However, the higher age of the study population could impact the event rates in diseases such as cardiovascular disease, where age is a risk factor.

5. Conclusion

Estimates of prevalence of HCU vary depending on factors such as the identification criteria, cohort definitions, and database or dataset used. Many patients with clinical presentations suggesting a diagnosis of HCU did not have an associated diagnosis of HCU, potentially indicating underdiagnosis or underreporting. Future research should explore alternative methods (such as the algorithm detailed in our analysis) to better diagnose and understand the true prevalence of HCU.

Funding

This work was funded by Travere Therapeutics, Inc. Travere Therapeutics, Inc. paid for the license to utilize Optum's de-identified Market Clarity Data.

Role of the funding source

The sponsor played a role in the study design, data interpretation, and decision to submit this article for publication. The sponsor also provided extensive review and comment on the writing of the manuscript.

CRediT authorship contribution statement

Mahim Jain: Writing – review & editing, Supervision, Conceptualization. **Mehul Shah:** Writing – review & editing, Methodology, Formal

analysis, Conceptualization. **Kamlesh M. Thakker:** Writing – review & editing, Investigation, Conceptualization. **Andrew Rava:** Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization. **Agness Pelts Block:** Writing – review & editing. **Colette Ndiba-Markey:** Writing – review & editing, Validation, Formal analysis. **Lionel Pinto:** Writing – review & editing, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Mahim Jain reports financial support was provided by Travere Therapeutics Inc. Kamlesh M. Thakker reports financial support was provided by Travere Therapeutics Inc. Mehul Shah reports a relationship with Travere Therapeutics Inc. that includes: employment and equity or stocks. Agness Pelts Block reports a relationship with Travere Therapeutics Inc. that includes: employment and equity or stocks. Lionel Pinto reports a relationship with Travere Therapeutics Inc. that includes: employment and equity or stocks. Andrew Rava and Colette Ndiba-Markey are employees of Genesis Research Group, which receives consulting fees from Travere Therapeutics, Inc. for this study and other related studies. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

M.J. has received consultancy fees from Travere Therapeutics, Inc.

M.S. is an employee and stock/options holder (vested and unvested) of Travere Therapeutics, Inc.

K.M.T has a consulting contract with Travere Therapeutics, Inc. and does not have any equity interest in Travere Therapeutics, Inc.

A.R. is an employee of Genesis Research Group, which receives consulting fees from Travere Therapeutics, Inc. for this study and other related studies.

A.P.B. is an employee and stockholder of Travere Therapeutics Inc.

C.N.M. is an employee of Genesis Research Group, which receives consulting fees from Travere Therapeutics, Inc. for this study and other related studies.

L.P. is a former employee and stock/options holder (vested and unvested) of Travere Therapeutics, Inc.

Data availability

The data that support the findings of this study are available from Optum, 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 Optum.

Acknowledgements

Medical writing support was provided by Jacqueline Michel of Genesis Research Group (Hoboken, NJ, USA), which was funded by Travere Therapeutics, Inc. Jihaeng Heo (Genesis Research Group at the time of the study), Tom Wasser (Genesis Research Group), and Wu Gong (Travere Therapeutics, Inc.) provided additional support with design and conduct of this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymgmr.2024.101101.

References

[1] S.H. Mudd, F. Skovby, H.L. Levy, et al., The natural history of homocystinuria due to cystathionine beta-synthase deficiency, Am. J. Hum. Genet. 37 (1) (Jan 1985) 1–31.
[2] Online Mendelian Inheritance in Man OMIM®. Johns Hopkins University, Baltimore, MD. MIM Number: 236200. Updated 08/03/2016, Accessed 05/25/23, https://www.omim.org/entry/236200, 2024.
[3] S.J. Sacharow, J.D. Picker, H.L. Levy, Homocystinuria caused by cystathionine beta-synthase deficiency, in: M.P. Adam, G.M. Mirzaa, R.A. Pagon, et al. (Eds.), GeneReviews(®), University of Washington, Seattle, 2024. Copyright © 1993–2023.
[4] A. Gerrard, C. Dawson, Homocystinuria diagnosis and management: it is not all classical, J. Clin. Pathol. 75 (11) (2022) 744–750, https://doi.org/10.1136/jcp-2021-208029.
[5] B.A. Maron, J. Loscalzo, The treatment of hyperhomocysteinemia, Annu. Rev. Med. 60 (2009) 39–54, https://doi.org/10.1146/annurev.med.60.041807.123308.
[6] A.A.M. Morris, V. Kožich, S. Santra, et al., Guidelines for the diagnosis and management of cystathionine beta-synthase deficiency, J. Inherit. Metab. Dis. 40 (1) (2017) 49–74, https://doi.org/10.1007/s10545-016-9979-0.
[7] M. Sellos-Moura, F. Glavin, D. Lapidus, K.A. Evans, L. Palmer, D.E. Irwin, Estimated prevalence of moderate to severely elevated total homocysteine levels in the United States: a missed opportunity for diagnosis of homocystinuria? Mol. Genet. Metab. 130 (1) (2020) 36–40, https://doi.org/10.1016/j.ymgme.2020.02.001.
[8] S. Yap, H. Rushe, P.M. Howard, E.R. Naughten, The intellectual abilities of early-treated individuals with pyridoxine-nonresponsive homocystinuria due to cystathionine beta-synthase deficiency, J. Inherit. Metab. Dis. 24 (4) (Aug 2001) 437–447, https://doi.org/10.1023/a:1010525528842.
[9] H. Gan-Schreier, M. Kebbewar, J. Fang-Hoffmann, et al., Newborn population screening for classic homocystinuria by determination of total homocysteine from Guthrie cards, J. Pediatr. 156 (3) (Mar 2010) 427–432, https://doi.org/10.1016/j.jpeds.2009.09.054.
[10] P. Mandava, T. Kent, Homocystinuria/Homocysteinemia 27, Medscape, 2018. Updated July. Accessed June 2, 2023, https://emedicine.medscape.com/article/1952251-overview#showall.
[11] M. Sellos-Moura, F. Glavin, D. Lapidus, K. Evans, C.R. Lew, D.E. Irwin, Prevalence, characteristics, and costs of diagnosed homocystinuria, elevated homocysteine, and phenylketonuria in the United States: a retrospective claims-based comparison, BMC Health Serv. Res. 20 (1) (Mar 6 2020) 183, https://doi.org/10.1186/s12913-020-5054-5.
[12] F. Skovby, M. Gaustadnes, S.H. Mudd, A revisit to the natural history of homocystinuria due to cystathionine beta-synthase deficiency, Mol. Genet. Metab. 99 (1) (Jan 2010) 1–3, https://doi.org/10.1016/j.ymgme.2009.09.009.
[13] K. Chapman, D. Bartke, M. McGlynn, H. Levy, C. Ficicioglu, Cystathionine beta synthase deficiency patients are being missed by newborn screening, Abstract. Genet. Med. 24 (3) (2022) S5, https://doi.org/10.1016/j.gim.2022.01.046.
[14] H.L. Levy, Early development of newborn screening for HCU and current challenges, Int. J. Neonatal. Screen. 7 (4) (Oct 25 2021), https://doi.org/10.3390/ijns7040067.
[15] C.I. Kaye, the Committee on Genetics, Newborn screening fact sheets, Pediatrics 118 (3) (2006) e934–e963, https://doi.org/10.1542/peds.2006-1783.
[16] J.H. Walter, N. Jahnke, T. Remmington, Newborn screening for homocystinuria, Cochrane Database Syst. Rev. (10) (2015) doi:10.1002/14651858.CD008840.pub4.
[17] C.A. Pickens, E. Courtney, S.L. Isenberg, C. Cuthbert, K. Petritis, Multiplexing homocysteine into first-tier newborn screening mass spectrometry assays using selective thiol derivatization, Clin. Chem. 69 (5) (2023) 470–481, https://doi.org/10.1093/clinchem/hvad007.
[18] H. Refsum, A.D. Smith, P.M. Ueland, et al., Facts and recommendations about Total homocysteine determinations: an expert opinion, Clin. Chem. 50 (1) (2004) 3–32, https://doi.org/10.1373/clinchem.2003.021634.
[19] M. Jain, L. Pinto, M. Shah, et al., Validation of a Patient Identification Algorithm to Estimate the Prevalence of Classical Homocystinuria (HCU) in the United States (US). Poster presented at: The Society for the Study of Inborn Errors of Metabolism (SSIEM) Annual Symposium, 2023. August 29–September 1, 2023; Jerusalem, Israel.