Poster
RWD50

Validation of a Claims Algorithm to Identify Patients With Myotonic Dystrophy Type 1 (DM1)

Yiting Wang, PhD[1]; Ankita Misra, MS, MPH[2]; Yaohua Zhang, PhD[1]; Fares Nigim, MD, PhD[1]; Tara MacCannell, PhD[1]

1. Vertex Pharmaceuticals Incorporated, Boston, Massachusetts, USA; 2. Optum Life Sciences, Eden Prairie, Minnesota, USA

INTRODUCTION

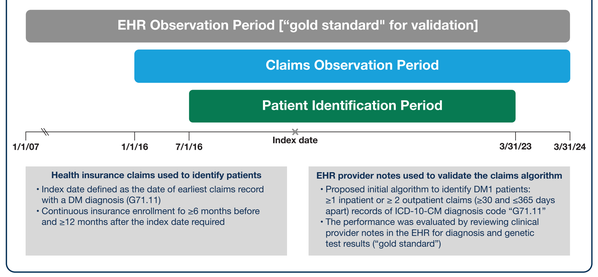
- Myotonic dystrophy type 1 (DM1) is an inherited multi-system disorder characterized by progressive muscle weakness and myotonia (difficulty relaxing muscles after contraction)
- In the US, there are no algorithms to specifically identify DM1 from insurance claims databases
- A commonly used, simple algorithm to identify other conditions from insurance claims databases: "≥ 1 inpatient or ≥ 2 outpatient claims" [1-4]
- A simple claims algorithm was adopted to identify DM1 and validate its performance against electronic clinical provider notes review using the Optum Market Clarity®, a linked claims-electronic health records (EHR) database in the US including 76 million individuals
- The goal of this study was to establish, for the first time, a validated algorithm for robust identification of DM1 patients from claims data to facilitate consistency across real world studies

OBJECTIVE

To validate the performance of a rule-based algorithm using ICD-10-CM codes

METHODS

Figure 1. Study Design



DM: myotonic dystrophy; **DM1:** myotonic dystrophy type 1; **EHR:** electronic health records

Study Population

- Claims population: Individuals with at least one record of ICD-10-CM code "G71.11" in patient identification period, with the date of the first record designated as the index date; continuous insurance enrollment of at least 6 months before and 12 months after the index date; age ≥10 years at index
- EHR validation sample: from the claims population, at least one provider note and DM-relevant key word; sufficient information to determine true DM1 vs not true DM1 (non-DM1) based on explicit diagnosis records or genetic test results in provider notes

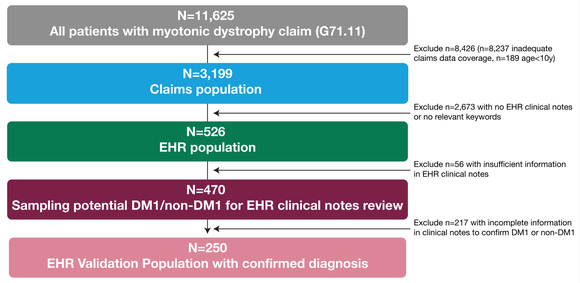
Analyses

- Positive predictive value (PPV): percent of individuals identified as having DM1 by the claims algorithm who were confirmed by EHR clinical notes review (gold standard)
- Sensitivity: percent of individuals with DM1 confirmed by EHR clinical notes review who were identified by the claims algorithm
- Differential sampling of electronic charts for review in the EHR validation sample was applied: 100% of potential non-DM1 and 87% of potential DM1 charts were differentially sampled into the EHR validation sample, based on keyword presence in EHR clinical notes. To account for the imbalanced sampling design, inverse sampling weights were applied in PPV estimation to restore source-population prevalence

Key Statistical Methods

All analyses were descriptive (i.e., mean, standard deviation [SD], counts, proportions) and no statistical hypothesis testing was performed. Summary statistics were reported overall, by EHR-validated true DM1, true non-DM1

Figure 2. Flowchart to Create the EHR Validation Sample



DM1: myotonic dystrophy type 1; **EHR:** electronic health record; **N:** number of individuals; **n:** size of subsample; **y:** years
N adjusted for sampling weights = 250; unadjusted N=253

Evaluation of the Algorithm

- 165 of 224 DM1 by claims algorithm were confirmed as true DM1 by EHR, with PPV 0.74

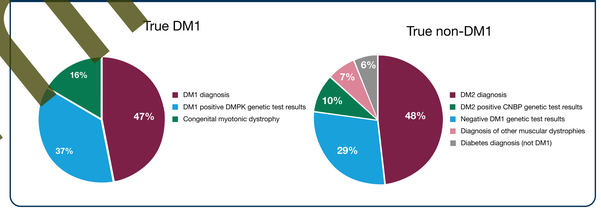
Table 1. Evaluation of the Algorithm

Number of patients[a]	EHR True DM1	EHR True non-DM1	Total
Claims algorithm DM1	165	59	224
Claims algorithm non-DM1	8	18	26
Total	173	77	250

DM1: myotonic dystrophy type 1; **EHR:** electronic health records
[a] Weighted counts to adjust for differential sampling of potential DM1 vs non-DM1 for charts review

- PPV = 165/224 = 0.74, 95% CI (0.68-0.79)
- Sensitivity = 165/173 = 0.95, 95% CI (0.92-0.99)
  - For claims algorithms, PPV> 0.7[1,5-7] is considered moderate to high quality, and high quality if sensitivity is also >0.7[8,9]

Figure 3. EHR Clinical Notes Review Details



CNBP: cellular nucleic acid-binding protein; **DM1:** myotonic dystrophy type 1; **DM2:** myotonic dystrophy type 2; **DMPK:** dystrophia myotonica protein kinase; **EHR:** electronic health record

Refined Algorithm Exploration

- A refined algorithm was built to further exclude non-DM1 patients
  - Further exclude patients with claims record of a stress test, or cellular nucleic acid-binding protein (CNBP) gene test, or aldolase test
- The tests above were selected from ~500 claims factors (symptoms, tests/procedures, medications) using a systematic selection process to identify factors that are specific to non-DM1 individuals

Table 2. Candidate Variables Identified to Refine the Initial Algorithm

	DM1	Non DM1	Std Diff >10%	Uncommon (<10%) in DM1, Common (≥10%) in non-DM1	Clinical decision
Stress test	5%	10%	17%	Yes	Rarely used in DM1
CNBP gene testing (for DM2 mutation)	4%	16%	42%	Yes	Specific DM2 gene test
Aldolase	3%	13%	39%	Yes	Rarely used in DM1

CNBP: cellular nucleic acid-binding protein; **DM1**: myotonic dystrophy type 1; **DM2**: myotonic dystrophy type 2; **Std**: standardized difference

- Refinement (to improve PPV) of the initial simple algorithm by additionally excluding 3 claims variables that met all three criteria on standardized difference, prevalence, and clinical input[10]

Table 3. Original Proposed Algorithm is Recommended

Claims algorithms	Definition of DM1 by health insurance claims	PPV 95% CI	Sensitivity 95% CI
Proposed	≥1 inpatient or ≥ 2 outpatient claims (≥30 and ≤365 days apart) records of ICD-10-CM diagnosis code "G71.11" for myotonic dystrophy	0.74 0.68-0.79	0.95 0.92-0.99
Refined (exploratory)	From the above, additionally exclude any of the test procedures: CNBP (DM2) test, stress test, or aldolase test	0.78 0.72-0.84	0.86 (0.81-0.91)
Sensitivity analysis 1	≥1 claims record with ICD-10-CM code "G71.11" for myotonic dystrophy	0.69 (0.63-0.75)	1
Sensitivity analysis 2	Initially proposed algorithm restricted to age <65 y	0.77 (0.72-0.83)	0.95 (0.92-0.98)

CI: confidence internval; **CNBP:** cellular nucleic acid-binding protein; **DM1:** myotonic dystrophy type 1; **DM2:** myotonic dystrophy type 2; **PPV:** positive predictive value; **y:** years

- Refined algorithm numerically improved PPV from 0.74 to 0.78
- Sensitivity analysis 1 showed that PPV<0.69
- The original proposed algorithm is recommended because the refined version has:
  - added complexity with little gain in PPV
  - considerable drop in sensitivity
- When restricted to patients aged <65 at index date, the PPV (95% CI) of the initial algorithm is 0.77 (0.72-0.83), and sensitivity (95%CI) 0.95 (0.92-0.98)

Table 4. Description of the EHR Validation Sample (Unweighted Counts)

EHR validation sample	Total (N=253)	True DM1 by EHR (N=170)	True non-DM1 by EHR (N=83)
Age, mean (SD)	43.77 (16.57)	39.81 (16.07)	51.89 (14.57)
Age group, years, n (%)			
10-<18	23 (9)	21 (12)	2 (2)
18-<50	121 (48)	93 (55)	28 (34)
50-<65	88 (35)	50 (29)	38 (46)
65+	21 (8)	6 (4)	15 (18)
Gender, n (%)			
Female	133 (52.57)	89 (52.35)	44 (53.01)
Male	120 (47.43)	81 (47.65)	39 (46.99)
Race/Ethnicity, n (%)			
Non-Hispanic White	170 (67.19)	113 (66.47)	57 (68.67)
Non-Hispanic Black	7 (2.77)	6 (3.53)	1 (1.20)
Non-Hispanic Asian	1 (0.40)	1 (0.59)	0 (0)
Hispanic	14 (5.53)	9 (5.29)	5 (6.02)
Other/Unknown	61 (24.11)	41(24.12)	20 (24.10)
Insurance type, n (%)			
Commercial	130 (53.72)	83 (51.23)	47 (58.75)
Medicare	68 (28.10)	56 (34.57)	12 (15.00)
Medicaid	44 (18.18)	23 (14.20)	21 (26.25)

DM1: myotonic dystrophy type 1; **EHR:** electronic health record; **N:** number of individuals; **n:** size of subsample; **SD:** standard deviation

CONCLUSIONS

- **This is the first validated claims algorithm for identifying DM1 from US administrative claims databases**
- **This validated algorithm is easy to apply in practice and has a good PPV and sensitivity**
- **For the purpose of generating insights about disease burden and costs of DM1, this algorithm enables consistent and reliable application across US claims databases**

Strengths:

- **First study to have used a rigorous approach to validate the claims algorithm against EHR clinical notes to specifically identify DM1 in the US**
- **The proposed claims algorithm has a good PPV and is simple to use in US insurance claims databases**

Limitations:

- **EHR provider notes used for validation may not capture complete medical records, including genetic testing results**
- **Insurance claims do not capture enough clinical details; the PPV of the simple algorithm only improved slightly by incorporating the 3 procedures found among ~500 claims factors**

- **Future research may explore more complex models to predict and classify DM1 phenotype, but the simple algorithm would be easier to understand and apply consistently**

Refrences

1. Kuang A, Xu C, Southern D, Sandhu N, Quan H. J Epidemiol Popul Health. *Validated administrative data-based ICD-10 algorithms for chronic conditions: A systematic review*. 2024;72:202744
2. Djibo DA, Margulis AV, McMahill-Walraven CN, Saltus CW, Shuminski P, Kaye JA, et al. Pharmacoepidemiol Drug Saf. *Validation of an ICD-10 case-finding algorithm for endometrial cancer in US insurance claims*. 2024;33:e5690
3. Grosse SD, Green NS, Reeves SL. Pediatr Blood Cancer. *Administrative data identify sickle cell disease: A critical review of approaches in U.S. health services research*. 2020;67(12):e28703.
4. Klabunde CN, Potosky AL, Legeler JM, Warren JL. J Clin Epidemiol. *Development of a comorbidity index using physician claims data*. 2000;53(12):1258-67.
5. Tonelli M, Wiebe N, Fortin M, Guthrie B, Hemmelgarn BR, James MT, et al. BMC Med Inform Decis Mak. *Methods for identifying 30 chronic conditions: application to administrative data*. 2015:15:31.
6. Tenny S, Hoffman MR. StatPearls. *Prevalence*. https://www.ncbi.nlm.nih.gov/books/NBK430867/ Accessed: 13 April 2026.
7. Chomistek AK, Franklin JM, Sobel RE, Marcus AF, Sinnott S, et al. Pharmacoepidemiol Drug Saf. *Development and Validation of Claims-Based Algorithms for Conjunctivitis and Keratitis*. 2024;33(11):e70052.
8. Vicente E, Ruiz de Sabando A, Garcia F, Gaston I, Ardanaz E, Ramos-Arroyo MA. Orphanet J Rare Dis. *Validation of diagnostic codes and epidemiologic trends of Huntington disease: a population-based study in Navarre, Spain*. 2021;16:77.
9. Chen X, Agiro A, Martin AS, Lucas AM, Haynes K. BMC Med Res Methodol. *Chart validation of an algorithm for identifying hereditary progressive muscular dystrophy in healthcare claims*. 2019;19(1):174.
10. Austin, P. C. (2009). *Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research*. Communications in Statistics - Simulation and Computation, 38(6), 1228–1234. https://doi.org/10.1080/03610910902859574

Author Disclosures

AM has nothing to disclose. **YW, FN, YZ** and **TM** are employees of Vertex Pharmaceuticals Incorporated and hold stock and/or stock options in the company.

Acknowledgments

Medical writing and editorial coordination was provided by Smitha Infante, PhD (SI), and graphic design support was provided by Alexandra Battaglia (AB). **SI** and **AB** are employees of Vertex Pharmaceuticals Incorporated and may hold stock and/or stock options at the company. Study VX24-DM1-008 is sponsored by Vertex Pharmaceuticals Incorporated.