

Development of an Algorithm to Identify Patients with Desmoid Tumors in a United States Claims Dataset

Chia Jie Tan¹, Connor Willis¹, Timothy Bell², Brad Tumminello^{2*}, Shengfan Zhou², Dipam Doshi², Carl Asche¹, Joe Biskuspiak¹, Diana Brixner¹, Anna Chalmers³, David Stenehjem^{1,4}

¹College of Pharmacy, University of Utah, Salt Lake City, UT, United States

²SpringWorks Therapeutics, Inc., Stamford, CT, United States

³Huntsman Cancer Institute, University of Utah Health, Salt Lake City, UT, United States

⁴College of Pharmacy, University of Minnesota, Duluth, MN, United States

*Former employee of SpringWorks Therapeutics, Inc.

CTOS 2025
Boca Raton, Florida
12–15 Nov 2025

Introduction

- Desmoid tumors (DT) are rare, locally aggressive soft-tissue tumors. Although DT do not metastasize, they can cause substantial symptom burden and potentially life-threatening organ damage.^{1,2}
- Prior to Oct 2023, there were no unique International Classification of Disease (ICD) diagnosis codes for DT.³ Instead, the diagnosis codes for neoplasms of uncertain behavior of connective and other soft tissue (ICD-9: 238.1 and ICD-10: D48.1) were generally used as the billing codes for DT.
- As both 238.1 and D48.1 were also used for other benign tumors, identification of patients with DT from claims databases prior to October 2023 had limited precision.
- To address this, we constructed an algorithm to identify patients who may have a diagnosis of DT in a United States claims dataset.

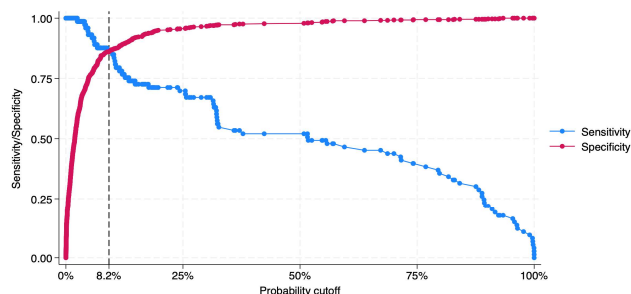
Methods

- Adult patients in the University of Utah Health network who had episodes of care tagged with ICD diagnosis codes 238.1 or D48.1 between 1 Jan 2011 and 31 Jul 2023 were identified.
- Patients with ICD diagnosis codes 238.1 or D48.1, new histopathological diagnosis of DT (in clinical notes and/or pathology reports) from 1 Jan 2011 to 31 Jul 2023 and ≥ 2 DT-related encounters were designated as cases.
- Patients with ICD diagnosis codes 238.1 or D48.1 but no histopathological diagnosis of DT served as controls.
- Cases and controls were randomly assigned to the model construction or validation dataset on a 1:1 split.
- To build the algorithm, stepwise logistic regression was performed using the model construction dataset:
 - Potential covariates: Demographics, ICD diagnosis codes and Current Procedural Terminology (CPT) procedure codes
 - Dependent variable: Confirmed diagnosis of DT
 - Potential covariates were sequentially added to the model and retained only if inclusion improved the adjusted R^2 of the model.
 - Odds ratios from the final model were applied to patient characteristics and associated diagnosis and procedure codes to estimate a predicted probability of DT.
 - Patients with a predicted probability exceeding a threshold were classified as having DT. The optimal threshold was determined by maximizing the product of sensitivity and specificity in the model construction dataset (Figure 1).⁴
- Model performance was assessed by evaluating the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) (Table 1) in the validation dataset.

Table 1. Definitions of model performance measures

Parameter	Definition
Sensitivity	The proportion of individuals with a confirmed DT diagnosis that were identified using the model above a specific % threshold as determined by the receiver operating curve
Specificity	The proportion of controls that were below a specific % threshold as determined by the receiver operating curve
Positive predictive value (PPV)	The likelihood that an individual above a specific % threshold as determined by the receiver operating curve has a confirmed DT diagnosis
Negative predictive value (NPV)	The likelihood that an individual below a specific % threshold as determined by the receiver operating curve does not have a confirmed DT diagnosis

Figure 1. Sensitivity and specificity across range of possible probability thresholds



Results

- Among 1460 patients with ICD diagnosis codes 238.1 or D48.1, 146 (10%) met criteria for cases while 1314 (90%) did not have DT and met criteria for controls. 73 cases and 657 controls (case-to-control ratio of 1:9) were each allocated to the construction and validation datasets.
- The final algorithm consisted of 12 variables (Table 2), including age in years, sex, ICD codes for neoplasms and CPT codes for biopsy and imaging procedures. The optimal threshold of predicted probability to classify a patient as having DT was at 8.2% (Figure 1).
- In the validation dataset:
 - The algorithm successfully identified 83.6% of patients with DT (sensitivity) and 84.0% of patients without DT (specificity).
 - Among patients classified as having DT by the algorithm, the probability of a confirmed diagnosis of DT was 36.8% (PPV).
 - The probability of a correct prediction among patients classified as not having DT by the algorithm was 97.9% (NPV).

Table 2. Variables in algorithm to identify patients with DT in claims datasets

Variable	Odds ratio
Age (in years)	0.97
Sex (Male vs female)	0.36
ICD code D13.2/D13.3/211.2 (Benign neoplasm of small bowel)	27.4
ICD code D48.3/D48.4/235/4 (Neoplasm of uncertain/unknown behavior of retroperitoneum/peritoneum)	9.6
ICD code D48.6/238.3 (Neoplasm of uncertain/unknown behavior of breast)	62.8
ICD code M72.2/M72.1/M72.4/728.7 (Other fibromatoses, pseudosarcomatous fibromatosis)	7.0
ICD code M72.8 (Other fibroblastic disorders)	97.2
CPT code 20206 (Biopsy of muscle with percutaneous needle)	13.0
CPT code 70553 (MRI brain and brain stem without contrast)	0.04
CPT code 71250 (Diagnostic CT thorax without contrast)	0.14
CPT code 71552 (MRI chest without and with contrast)	53.8
CPT code 74183 (MRI abdomen without and with contrast)	4.5
Constant	0.30
Optimal threshold for identifying patient with DT	8.2%

Discussion and Conclusion

- The main challenge to building an algorithm to identify patients with DT was the low incidence of DT, which led to a small number of true positives relative to false positives regardless of the algorithm sensitivity. To mitigate this, the algorithm was targeted to patients with billing codes 238.1 or D48.1.
- The predictive algorithm more accurately identified patients with DT compared to using the non-specific ICD-10 codes 238.1 and D48.1 alone. In our sample of patients who had ICD diagnosis codes 238.1 or D48.1 and were included for algorithm development, only 10% had a histopathologic diagnosis of DT. Using the algorithm we developed, the likelihood that an individual identified as having DT was accurate increased to 36.8% (PPV).
- Using the algorithm may further improve the precision of identifying patients with DT in claims databases prior to the broad adoption of DT-specific ICD diagnosis codes, allowing longitudinal epidemiological and real-world data studies with observation periods that extend to before Oct 2023.

Acknowledgements

This study was supported by a research agreement with SpringWorks Therapeutics, Inc.

References

- Kasper B, et al. *Oncologist*. 2011. 16(5):682-93.
- Husson O, et al. *Support Care Cancer*. 2019. 27(3):965-80.
- Desmoid Tumor Research Foundation. Ten New ICD-10-CM Diagnosis Codes For Desmoid Tumors Go Into Effect on October 1, 2023, in the United States. [Press release] Available at: <https://dtrf.org/about-dtrf/news-updates/ten-new-icd-10-cm-diagnosis-codes-for-desmoid-tumors-go-into-effect-on-october-1-2023-in-the-united-states/>. Accessed 15 Sep 2025.
- Liu X. *Stat Med*. 2012. 31(23):2676-86.

Disclosures

CJT, CW, CA, JB, DB, DS received research funding from SpringWorks Therapeutics, Inc. to conduct this study. TB, BT, SZ, DD were employees of SpringWorks Therapeutics, Inc. while AC was on the speaker's bureau of SpringWorks Therapeutics, Inc. at the time of the study.