# IDENTIFYING TENOSYNOVIAL GIANT CELL TUMOR (TGCT) IN SECONDARY DATA: METHODOLOGICAL CHALLENGES, EMERGING OPPORTUNITIES, AND FUTURE DIRECTIONS

SA29

Makoto Endo,[1] Dina Oksen,[2] Emmanuelle Boutmy,[3] Roberto Sichera,[4] Uladzislau Yanuts,[4] Doreen A Kahangire[3]

[1]Kyushu University, Fukuoka, Japan; [2]Merck Biopharma Co., Ltd, Tokyo, Japan; [3]The Healthcare business of Merck KGaA, Darmstadt, Germany; [4]Valos Srl, Genoa, Italy

## CONCLUSION

- **The algorithm identified patients with TGCT who had demographic, clinical, and treatment characteristics that were consistent with prior publications,[1,2] suggesting that the identification strategy is clinically substantiated and plausible**
- **Standardizing diagnostic criteria and International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) coding practices across geographies is crucial for improving global TGCT surveillance, particularly as TGCT is a biologically driven colony-stimulating factor-1 (CSF-1) disease that is clinically actionable with highly specific CSF-1 receptor systemic therapies[3]**
- **Integration with complementary data sources, such as pathology, radiologic imaging, and natural language processing of electronic health records, will be essential for accurate identification**
- **This study adds to the limited body of real-world publications about TGCT**
- **Further algorithm development may include clinical validation with histologic confirmation to determine further sensitivity and specificity, as well as evaluation across additional independent datasets**
- **The algorithm may assist healthcare providers and data systems in applying TGCT-related ICD-10 codes more consistently, thereby improving data quality for research, surveillance, and patient outcomes in this underserved population**

## PLAIN LANGUAGE SUMMARY

- **TGCT is a rare tumor that can affect any joint in relatively young adults**
- **Large insurance and hospital databases can show how people with TGCT are diagnosed and treated, but TGCT is often hard to recognize, miscoded, or placed under broad labels, which makes it difficult to identify patients**
- **As new TGCT treatments emerge, reliable ways to identify patients with TGCT in real-world data are needed so that care and research are based on accurate, patient-relevant information**
- **We tested methods to find TGCT cases in US and Japanese insurance data using diagnosis and procedure codes, prescriptions, specialist visits, and treatment patterns**
- **Early results were promising but limited by coding differences, variable documentation, and lack of biopsy data**
- **Standardizing TGCT diagnosis and coding, and linking claims with other data sources (e.g., pathology, imaging, electronic health records), will be key to tracking this rare disease and supporting better research and policy**

## INTRODUCTION

### TENOSYNOVIAL GIANT CELL TUMOR

- TGCT is a soft tissue tumor that can affect the synovium (joint lining), bursae, or tendon sheaths of any joint (most commonly the knee) and cause substantial symptom burden, joint destruction, and life-long functional limitations (**Figure 1**)[3–5]
- The estimated annual global incidence of TGCT is 43 cases per 1 million person-years,[6] and it is usually diagnosed in young, otherwise healthy adults (median age 38 years)[3–5]
- The two most common subtypes of TGCT are localized (L-TGCT; also known as nodular) and diffuse (D-TGCT)[3,4]
  - L-TGCT is more common and has a generally indolent course of disease[3,4]
  - D-TGCT represents 10–20% of cases and is characterized by extensive, infiltrative tumor growth[3,4]

### CHALLENGES WITH IDENTIFYING TGCT

- Due to its rarity and similarity of symptoms (i.e., pain, swelling, stiffness) with other common joint conditions, such as arthritis, TGCT can be challenging to diagnose and identify[4,5]
- Real-world evidence from administrative claims data can help to characterize TGCT, but challenges, such as non-specific coding, further complicate the identification of this patient population

### IMPROVING TGCT SURVEILLANCE

- There is an important need for methodologies to identify patients with TGCT within existing real-world datasets[7–9]
- A validated algorithm to improve identification of patients with TGCT would lead to more robust data that can help to address critical issues like diagnostic ambiguity, misclassification, and coding problems, as well as better clinical decision-making and long-term outcomes for patients

**Figure 1. Frequently reported symptoms of TGCT[4]**

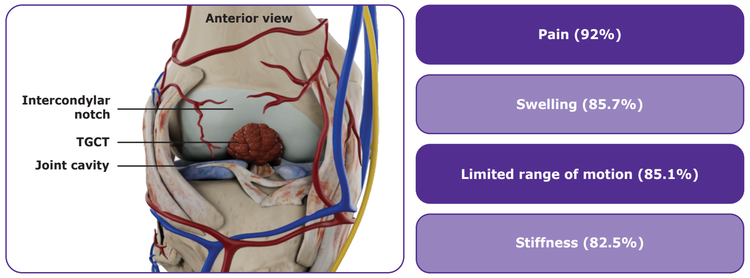



Image used under license from TurboSquid by Shutterstock

## OBJECTIVE

- To develop an algorithm that is verified based on scalability, reproducibility, and subtype identification to accurately identify patients with TGCT

## METHODS

### REAL-WORLD DATABASES

- This observational cohort study utilized de-identified patient-level data abstracted from two nationwide administrative claims databases (**Figure 2**)
  - Merative™ MarketScan® Research Databases (US) and JMDC Claims Database (Japan) include longitudinal, in-depth data for medical, prescription, and healthcare expenditures

### ALGORITHM ATTRIBUTES

- ICD-10 codes were used to identify patients with TGCT
  - ICD-10 Clinical Modification (CM) in the US-based study
  - ICD-10 in the Japan-based study
- TGCT case definition was operationalized using specific ICD-10 codes for L-TGCT and D-TGCT while excluding other codes (e.g., desmoid tumor) (**Table 1**)
- Orthopedic surgeons provided clinical input to ensure correct patient identification and differentiation between TGCT subtypes
- Inclusion criteria included having TGCT case definition, ⩾1 inpatient (IP)/emergency department or ⩾1 outpatient (OP) claim, and/or prescription for approved systemic therapy for TGCT (e.g., pexidartinib, vimseltinib)
- Index date was defined as the date of the first TGCT-related medical claim

**Figure 2. Study design**

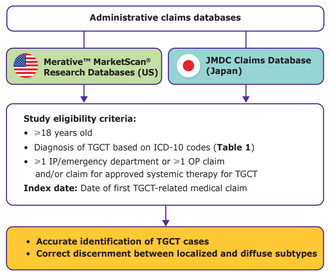



**Table 1. Specific ICD-10 codes for diagnosis of TGCT[a]**

| TGCT subtype | Inclusion criteria | | Exclusion criteria (examples[b]) | |
|---|---|---|---|---|
| | Codes | Description | Codes | Description |
| **Localized** | • D48.1 + D21.X | • Neoplasm of uncertain behavior of connective and other soft tissue<br>• Other benign neoplasms of connective and other soft tissue | • D48.11<br>• D48.2<br>• D21 | • Desmoid tumor<br>• Neoplasm of uncertain behavior of peripheral nerves and autonomic nervous system<br>• Other benign neoplasms of connective and other soft tissue |
| **Diffuse** | • M12.2<br>• D48.1 + M12.2X<br>• D21.X + M12.2X<br>• D21.X + D48.1 + M12.2X | • Other and unspecified arthropathy<br>• Neoplasm of uncertain behavior of connective and other soft tissue<br>• Villonodular synovitis (pigmented) | | |

[a]Selection of ICD codes was verified by clinician experts in TGCT; [b]In total, at least 30 ICD-10 codes were excluded

### ALGORITHM PERFORMANCE

- Algorithm performance was assessed based on scalability, reproducibility, and subtype identification
- Scalability
  - The algorithm was applied consistently across large US- and Japan-based datasets without substantial manual adaptation
  - This ensured stable performance despite differences in coding practices and healthcare utilization patterns
- Reproducibility
  - Reproducibility was evaluated by applying the algorithm independently across the two datasets (US and Japan) and confirming consistent TGCT case identification
  - For example, the algorithm was evaluated for consistent TGCT case identification when applied across independent datasets with repeated runs
  - This demonstrated stability of outputs under varying cohort sizes and time windows
- Subtype identification
  - Correctly distinguishing between L-TGCT and D-TGCT based on clinician-verified code combinations, ensuring accurate stratification aligned with real-world diagnostic pathways

## RESULTS

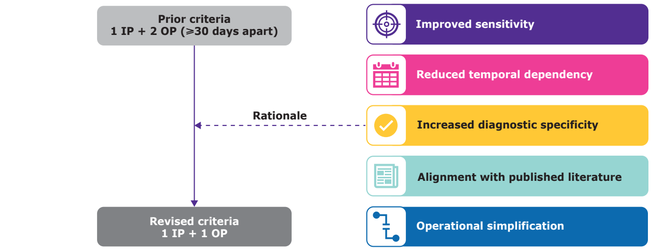
### PRELIMINARY ANALYSES

- The algorithm leveraged ICD-10 diagnosis and procedure codes, World Health Organization (WHO)-Anatomical Therapeutic Chemical codes, specialist visits, and treatment codes
- Preliminary analyses supported credibility and clinical plausibility of the identification strategy, indicating that administrative healthcare claims-based algorithms can enhance TGCT case ascertainment regardless of coding system, geography, healthcare setting, and study population

### REVISION OF CLAIMS CRITERION

- Based on clinical feedback, the inclusion criterion for claims was amended from 1 IP + 2 OP (⩾30 days apart) to 1 IP + 1 OP for the following reasons (**Figure 3**):
  - Improved sensitivity: the 2-visit OP with a 30-day separation rule underestimated true cases, particularly among patients with limited healthcare access or short insurance enrollment windows
  - Reduced temporal dependency: the 30-day separation rule introduced look-back period constraints and excluded incident cases with insufficient follow-up time
  - Diagnostic specificity of contemporary coding: with ICD-10 coding granularity, a single OP claim with a specific diagnosis code carries sufficient diagnostic confidence, comparable with the certainty attributed to a single IP claim
  - Alignment with published literature: published studies indicated that single-claim OP algorithms maintain acceptable positive predictive value when the diagnosis code is highly specific for the condition of interest[10]
  - Operational simplification: reducing the OP requirement to a single claim simplified algorithm implementation and index date determination without materially compromising case specificity
- Compared with the original claims criterion, the revised claims criterion allowed for greater flexibility and tailoring for TGCT, which can be challenging to diagnose, and for higher algorithm sensitivity

**Figure 3. Rationale for update of claims criterion**




### CHALLENGES

- Challenges with developing the algorithm included:
  - Diagnostic ambiguity due to overlapping conditions
  - Documentation variability across providers and healthcare settings
  - Absence of histopathologic confirmation
- These limitations have been reported with similar studies and are inherent to retrospective studies utilizing administrative claims databases[1,8–10]

### ICD SYSTEM CONSTRAINTS

- Heterogeneity of ICD subtype coding and lack of consistency across ICD coding systems used across geographies[2,8,9]
- In addition, the WHO stopped maintaining ICD-10 in 2018 and future enhancements will be introduced only in ICD-11[11]
- Countries have also had delayed adoption and have inconsistent implementation of ICD codes[11,12]

### REFERENCES

1. Dharmani C et al. Future Oncol 2024;20:1079–97
2. de Luise C et al. Pharmacoepidemiol Drug Saf 2021;30:1153–61
3. Stacchiotti S et al. Cancer Treat Rev 2023;112:102491
4. Stern S et al. Future Oncol 2025;21:1501–10 (and Suppl.)
5. Bernthal NM et al. Orphanet J Rare Dis 2021;16:191
6. Mastboom MJL et al. Acta Orthop 2017;88:688–94
7. Ehrenstein V et al. J Rheumatol 2017;44:1476–83
8. Burton TM et al. Clin Ther 2018;40:593–602.e1
9. Hess LM et al. Perspect Health Inf Manag 2019;16(Spring):1a
10. Rector TS et al. Health Serv Res 2004;39:1839–58
11. World Health Organization. https://www.who.int/standards/classifications/frequently-asked-questions/icd-11-implementation [Accessed March 26, 2026]
12. Imai T. (2022). Nakashima, N. (eds). Epidemiologic Research on Real-World Medical Data in Japan. https://doi.org/10.1007/978-981-19-1622-9_5 [Accessed March 26, 2026]

### ACKNOWLEDGMENTS

Medical writing support, under the direction of the authors, was provided by Trang H Au, PharmD, of the Publications Division of Omnicom Health Medical Communications, funded by the Healthcare business of Merck KGaA, Darmstadt, Germany in accordance with Good Publication Practice (GPP 2022) guidelines

This study was sponsored by the Healthcare business of Merck KGaA, Darmstadt, Germany (CrossRef Funder ID: 10.13039/100009945)

### AUTHOR DISCLOSURES

Makoto Endo, Uladzislau Yanuts, and Roberto Sichera have no conflicts of interest to disclose. Emmanuelle Boutmy and Doreen Kahangire are currently employed by Merck KGaA, Darmstadt, Germany and its affiliates. Dina Oksen was previously employed by Merck KGaA, Darmstadt, Germany and its affiliates.



**Poster code SA29. Presented at The Professional Society for Health Economics and Outcomes Research (ISPOR) 2026 Congress, 17–20 May 2026, Philadelphia, PA, USA**