In medicine, ontology refers to the formal representation of medical concepts to enable clear and unambiguous communication through defined medical terms, often using the International Classification of Diseases (ICD) codes. The development of medical ontologies is rigorous and lengthy, involving the collaboration of knowledge engineers and physicians1 . Despite these efforts, medical ontologies often lack specificity for a particular disease and are influenced by administrative processes2,3 . Recent advancements in natural language processing and the availability of large amounts of semantically richmedicalinformation have given rise to thefield of ontology learning. Ontology learning involves the semi-automatic identification and extraction of key concepts from text to enable the construction of an ontology4 , and can increase the accuracy of disease representations5 . Real-world data is now widely used to assess disease burden, treatment effectiveness, and safety. Unlike controlled clinical trials, researchers lack control and guidance over data reporting in real-world studies. Furthermore, real-world evidence studies often suffer from a lack of transparency regarding the reliability of selected ontologies, and a more rigorous assessment of the ontologies utilized is warranted. Here, we introduce the DiseaseSpecific Medical Ontology Learning (DiSMOL) framework, an innovative approach to ontology learning where pre-existing ontologies are enriched with real-world data and carefully evaluated. DiSMOL uses physician notes to extend medical ontologies, providing first-hand access to real-world reporting practices and language use among healthcare providers. This approach involves elements of deep phenotyping, particularly in systematically characterizing and learning about disease symptoms. However, it is also crucial to recognize that our work extends beyond mere symptom documentation. By incorporating these symptoms into an existing knowledge base, we essentially engage in a form of ontology learning, expanding and refining this knowledge base, which already has a specific structural framework. To demonstrate the potential of DiSMOL, we applied theframework to examine the burden of daytime impairment in people with insomnia disorder, the most prevalent worldwide sleeping disorder. Daytime impairment is a required feature of insomnia diagnosis6 and a very relevant aspect to be addressed for individuals affected by the disease. Improving daytime functioning has been identified as the most important attribute of an insomnia treatment by people with insomnia disorder7 . Despite its significance and importance for individuals with insomnia, symptoms describing daytime impairment are rarely coded. Therefore, DiSMOL aims to identify as many genuine daytime impairment symptoms as possible while minimizing the inclusion of spurious symptoms to provide an accurate representation of this critical aspect of insomnia disorder. In this paper, we show that DiSMOL is able to detect symptoms of daytime impairment in people with insomnia with greater sensitivity compared to ICD-10 codes. Additionally, the framework reveals significant increases in daytime impairment symptoms following benzodiazepine use (18.9%), while traditional ICD codes do not detect any significant change.