Real-World Endpoint Mapping and Identification of Evolving Phenotypes of Pompe Disease using Machine Learning

RWD22

Edin Guso, M.Sc.[1], Atef Zaher, MD[2], Kenneth I. Berger, MD[3], Vassia Liakoni, PhD[1], Martin Beaussart, MSc[1], Rory Burke, MSc[1], David Castellano Falcon, MSc[1], Vahid Esmaeili, PhD[1], Leon van Wouwe, MSc[1], Christopher Michael Rudolf[1], Andres Rondon[4], Kristina An Haack, MD[4]

[1]Volv Global, Épalinges, Switzerland, [2]Sanofi, Toronto, ON, Canada, [3]Sanofi, Morristown, NJ, USA, [4]Sanofi, Gentilly, France.

SCAN ME to download a copy of this poster!
Copies of this poster obtained through QR code are for personal use only.

INTRODUCTION

- Pompe disease (PD) is a rare, genetic disease caused by deficiency of the enzyme acid alpha-glucosidase, which leads to glycogen accumulation in muscle cells resulting in progressive and irreversible damage to skeletal, respiratory, cardiac, and smooth muscles.[1]
- As enzyme replacement therapy (ERT) extends survival, patients may experience new long-term manifestations that are not fully assessed in clinical trials and/or routine clinical practice.
- Claims data comprise a large volume of real-world data (RWD) for analysis, however, diagnosis and procedure codes do not capture endpoints assessed in clinical trials.

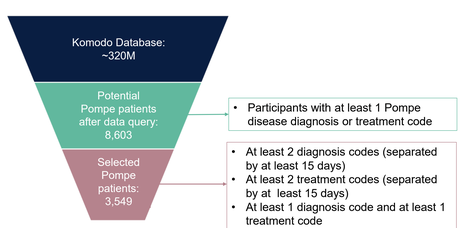
OBJECTIVE

- To assess whether key PD-related clinical endpoints and emerging disease manifestations can be identified in real-world claims data using machine learning

METHODS

- Patients with PD (infantile and late onset) were selected in a large U.S. administrative claims database (Komodo Health) using confirmed diagnosis and/or treatment records between Jan 2016-Dec 2025 (**Figure 1**).
- Three approaches were used to define PD features and phenotypes:
  - Literature review
  - Predefined clinical endpoints from expert opinion
  - Data-driven approach based on claims data
- Machine learning models were developed to map predefined clinical endpoints to diagnosis, procedure, and treatment codes.
  - Codes were prioritized and selected according to their similarity to each endpoint
- Prevalence of mapped and newly discovered disease features was evaluated in the PD cohort and compared with a demographically (age and sex) matched control population (n=288,563)

Figure 1. Pompe cohort selection and quality criteria



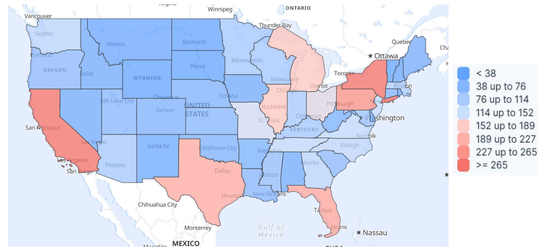
RESULTS

- A total of 3,549 patients with PD were selected.
  - Across the US, highest concentrations were in California (303), New York (293), Pennsylvania (226), Florida (200), Texas (196) (**Figure 2**).
  - The sex distribution was relatively similar between males (1650, 46.5%) and females (1899, 53.5%).
  - A large proportion of newborns (age at diagnosis<1; 794/3549, 22.4%) and children (1<age at diagnosis<5; 589/3549, 16.6%) in both sexes potentially reflects increased newborn screening utilization.
  - At diagnosis, male median age (10.4) appeared significantly younger than female median age (30.4).

CONCLUSIONS

1. Machine learning mapped 46 of 67 pre-defined Pompe disease clinical endpoints to claims-based RWD.
2. All mapped endpoints were significantly more prevalent in patients with PD than in controls.
3. Data-driven discovery identified additional distinctive features, including cardiovascular, respiratory, systemic, and healthcare-utilization manifestations.
4. These findings support the use of claims-based RWD and machine learning to identify clinically meaningful endpoints for future natural history studies and clinical trials.

Figure 2. Distribution of selected Pompe patients per State

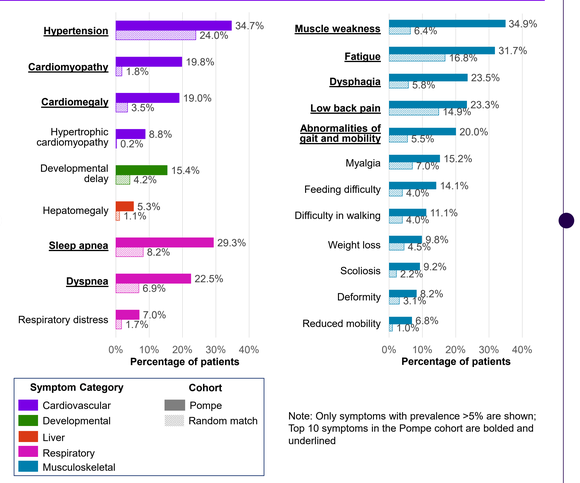


- Most frequently reported treatments include: alglucosidase alfa (30.1%), avalglucosidase alfa (16.3%), cipaglucosidase alfa (1.9%); the percentage of patients who ever received ERT during the study period was 32.9%.

Mapping Literature-based PD Symptoms to RWD

- The prevalence of all literature-based symptoms was significantly higher in the Pompe cohort compared to the random match cohort (p-value <0.01, likelihood ratio test), except for subarachnoid hemorrhage (**Figure 3**).
- The prevalence of the top 10 literature-based symptoms showed greater statistical significance between the Pompe cohort compared to the random match cohorts (p-value <0.001, likelihood ratio test), indicating a strong clinical association with disease (**Figure 3**).

Figure 3. Prevalence of literature-based symptoms in Pompe cohort compared to random match cohort

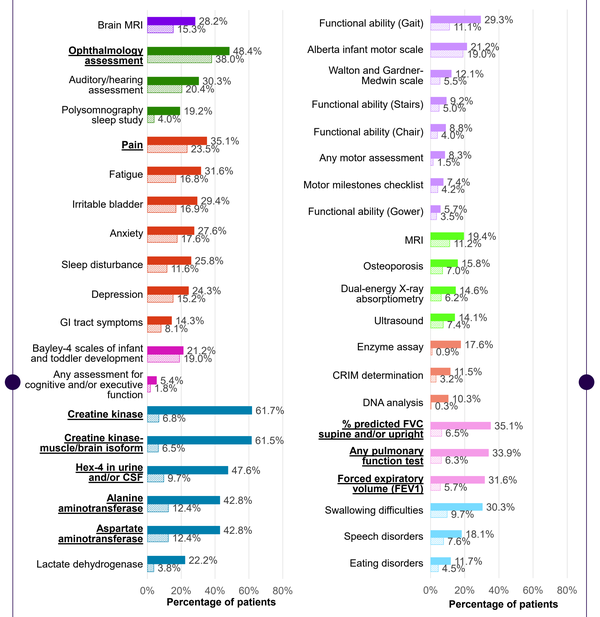


Note: Only symptoms with prevalence >5% are shown; Top 10 symptoms in the Pompe cohort are bolded and underlined

Mapping PD Clinical Endpoints to RWD

- Machine learning mapped 46 of the 67 predefined clinical endpoints to codes present in the claims-based RWD.
- The most relevant differences between the PD cohort and random matched cohort were detected in the following endpoint categories: lab biomarkers (74.8% vs 21.5%), pulmonary function test (35.3% vs 6.7%), Pompe diagnostic workup (29.7% vs 4.2%), speech lingual assessment (38.5% vs 16.3%).
- The most frequent individual endpoints mapped in the PD cohort were: creatine kinase (CK, 61.7%), creatine kinase-muscle/brain isoform (CK-MB, 61.5%), ophthalmology assessment (48.4%), Hex-4 (47.6%), alanine aminotransferase (42.8%), aspartate aminotransferase (42.8%), % predicted FVC supine and/or upright (35.1%), and pain (35.1%) (**Figure 4**).
- Significantly higher prevalence was observed in the PD cohort compared to controls (p-value <0.001, likelihood ratio test for all endpoints, except Bayley-4 scale and Alberta infant motor scale, which had p-values <0.01) (**Figure 4**).

Figure 4. Prevalence of predefined clinical endpoints in Pompe cohort compared to random match cohort

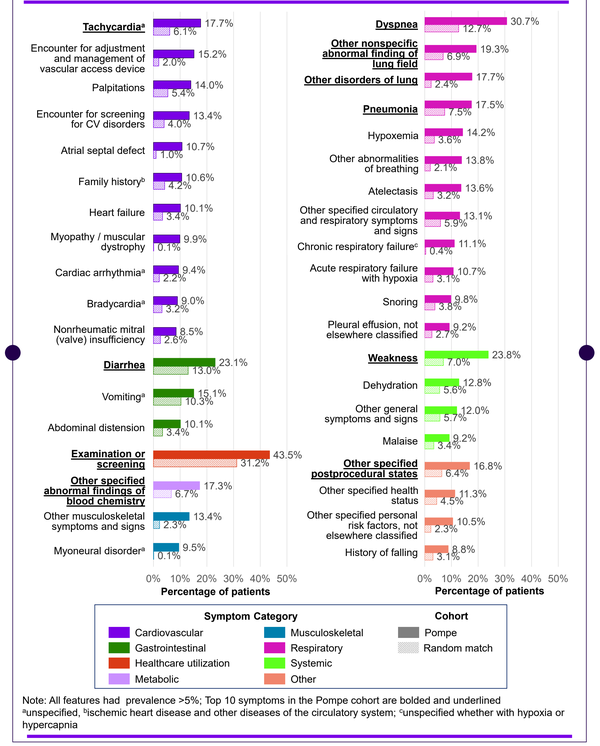


Note: Only symptoms with prevalence >5% are shown; Top 10 symptoms in the Pompe cohort are bolded and underlined
CSF, cerebrospinal fluid; CRIM, Cross-reactive immunological material; FVC, forced vital capacity; Hex-4, Hexose tetrasaccharide-4; MRI, magnetic resonance imaging

Unsupervised and Unbiased (Not-Prespecified) Discovery of PD Feature Categories

- Several feature categories were reported in significantly larger proportions in the PD cohort compared to random match cohorts (p-value <0.001, likelihood ratio test): cardiovascular (59.1% vs. 20.8%), respiratory (58.3% vs. 26.7%), healthcare utilization (43.5% vs 31.2%) and systemic (39.6% vs 16.5%).
- The most frequent data-driven discovered features in the PD cohort were: Examination or screening (43.5%), dyspnea (30.7%), weakness (23.8%), diarrhea (23.1%), and multiple respiratory abnormalities such as chronic respiratory failure, hypoxemia, and pneumonia (9–20%) (**Figure 5**).
- The prevalence of all data-driven discovery features was significantly higher in the PD cohort compared to random match cohorts (p-value <0.001, likelihood ratio test) (**Figure 5**).

Figure 5. Prevalence of data-driven endpoints in Pompe cohort compared to random match cohort



Note: All features had prevalence >5%; Top 10 symptoms in the Pompe cohort are bolded and underlined
[a]unspecified, [b]ischemic heart disease and other diseases of the circulatory system; [c]unspecified whether with hypoxia or hypercapnia

Study Limitations

- Claims data captures the occurrence of tests/diagnoses, not their results — we cannot assess endpoint values, only whether an endpoint-related event was recorded.
- The AI-driven code matching, while validated with manual clinical review, may include false positives.
- The cohort may include some misdiagnosed patients despite quality filtering.

The Professional Society for Health Economics and Outcomes Research (ISPOR), May 17-20, 2026

REFERENCES:

1. Reuser AJJ, et al. Pompe Disease: Glycogen Storage Disease Type II, Acid α-Glucosidase (Acid Maltase) Deficiency. In: Valle DL, Antonarakis S, Ballabio A, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. New York, NY: McGraw-Hill Education 2019.

ACKNOWLEDGMENTS:
Medical writing support for this poster was provided by Ronnie Yoo of Sanofi.

FUNDING: The study was funded by Sanofi.

DISCLOSURES:
EG, VL, MB, RB, DCF, VE, LvW, CMR — Employees of Volv Global; AZ, KIB, AR, KAH — Employees of Sanofi and may hold stock/stock options