Data Standardization of Claims Data in the InGef Research Database: A Comparison of Data Models for Epidemiological and Economic Studies of Rare Genetic Diseases in Germany

Josephine Jacob[1], Raeleesha Norris[1], Dominik Obermüller[1], Marco Alibone[1], Marion Ludwig[1]

[1]InGef – Institute for Applied Health Research Berlin GmbH | Otto-Ostrowski-Str. 5, 10249 Berlin, Germany | info@ingef.de

Conflict of interests

The work has not been funded by a third party but was exclusively funded by InGef. All authors have no relationships to disclose.

Background

- Research on rare genetic diseases is challenging due to their low prevalence and complex clinical manifestations.
- Common data models (CDM), such as the Observational Medical Outcomes Partnership (OMOP) model, play a crucial role in research, especially for exchanging data across countries or institutions.
- The OMOP CDM could be particularly helpful when investigating rare genetic diseases, where distributed analyses across different datasets and countries are essential to generate valid and robust evidence.

Objectives

- This study compares the prevalence and annual direct costs using anonymized claims data from the InGef research database (RDB) in its original source format and in the OMOP CDM format of the following selected rare genetic diseases
  - spinal muscular atrophy type I (SMA) (ICD 10 GM G12.0)
  - Huntington's disease (HD) (ICD 10 GM G10)
  - beta-thalassemia (BT) (ICD 10 GM D56.1)
  - cystic fibrosis (CF) (ICD 10 GM E84)
  - hereditary retinal dystrophy (HRD) (ICD 10 GM H35.5).

Methods

- Cross-sectional analysis for each disease in 2022.
- Data source: InGef database, consisting of approx. 10 million insured persons[1].
- Prevalence was directly standardized to the German population. Costs were analyzed as average total direct costs (i.e., the total costs of inpatient care, outpatient care, medication) per patient per year.

Results

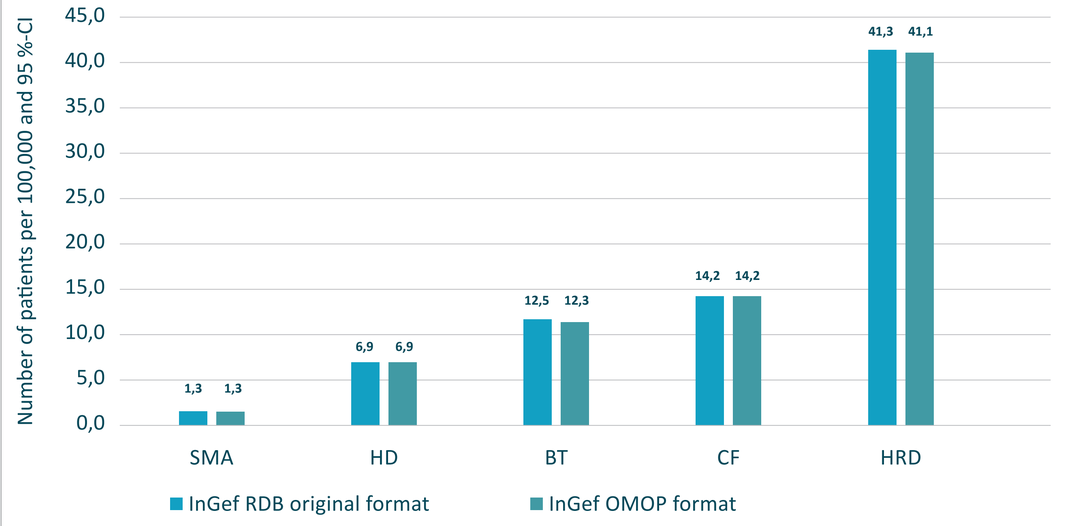
- Age and sex distribution of patient population of selected genetic rare diseases *(Table 1)* in 2022.

Tab. 1: Age and sex distribution of patients with rare genetic diseases by data source

Disease	InGef RDB original format			InGef OMOP format		
	n patients	female (%)	age (mean (SD))	n patients	female (%)	age (mean (SD))
HRD	2070	53.9	60.0 (20.3)	3055	53.9	60.0 (20.3)
CF	1216	47.9	19.8 (19.2	1216	47.9	19.8 (19.2)
BT	1043	56.6	42.9 (21.0)	1019	56.2	42.7 (21.0)
HD	578	52.9	57.1 (13.1)	576	53.2	57.2 (13.1)
SMa	110	46.6	30.3 (28.3)	109	45.9	29.9 (28.1)

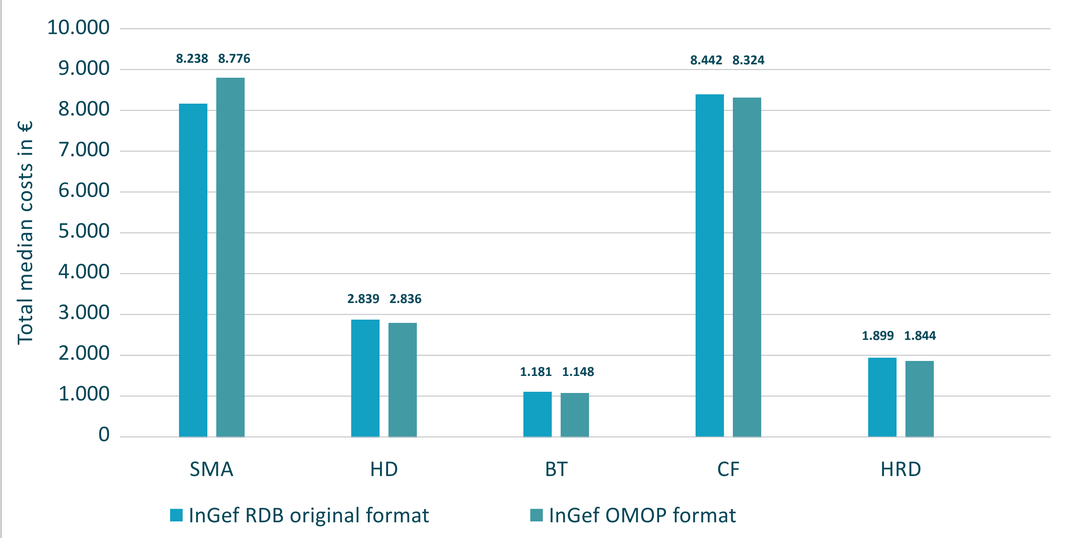
- Prevalences for selected genetic rare diseases in 2022 by data source *(Fig. 1)*.

Fig. 1: Prevalence of genetic rare diseases in InGef RDB and InGef OMOP database in 2022



- Total median costs in € for all diseases between the source and standardized data were: CF 8.442 vs. 8.324; HD 2.839 vs. 2.836; HRD 1.899 vs. 1.844; BT 1.181 vs. 1.148; SMA 8.238 vs. 8.776 *(Fig. 2)*.

Fig. 2: Total median costs in € by disease in InGef RDB and InGef OMOP database in 2022



Conclusion

- The standardized prevalence of rare genetic diseases and the associated median total costs were similar between InGef original and OMOP data.
- The observed minor discrepancies between source and OMOP results were mainly caused by differences in data preprocessing. Efforts to harmonize data preprocessing are ongoing.
- The InGef RDB, both in its source format and transformed into the OMOP CDM, is a valid database for studying rare diseases due to its size and longitudinal nature.

[1] Ludwig M, Enders D, Basedow F, Walker J, Jacob J. Sampling strategy, characteristics and representativeness of the InGef research database. Public Health. 2022 May;206:57–62