Examining the fitness-for-purpose of European real-world data sources for external comparators in haematological malignancies

Saunders A[1], Hogg C[1], Jeswani N[2], Masi D[1], Dushkin K[1], Papsch R[1]

1. IQVIA, London, UK, 2. RxPx, London, UK.

OBJECTIVE

- External comparators are useful to complement single-arm trials by providing real-world context on comparator treatments. They are increasingly of interest to regulators and payers in rare diseases and rare sub-populations of more common diseases.
- We aimed to understand the fitness-for-purpose of real-world data sources for haematology-oncology external comparators, with the intention of bringing together qualified data sources and research expertise to establish a research collaboration network.

METHODS

- Literature reviews, desk research, and recent/ongoing in-house clinical studies were reviewed to identify European data sources involved in real-world haematology-oncology data generation.
- Of 332 European data sources identified, outreach was conducted with data sources via detailed questionnaires as part of routine study activities to ascertain in-depth information on patient counts, variable availability, data quality, and operational aspects of data access.
- Patient counts reported by data source for specific indications and time periods were normalized to represent average prevalent patients per year, in patients-per-year, by dividing the number of prevalent patients reported for each indication by the number of years surveilled*. Because 9 of the 61 data sources did not report patient counts, the normalized patient counts were scaled up by a factor of 9/61 (14.8%) to approximate the prevalence per indication that could be expected across all 61 assessed data sources.

*Time frames surveilled ranged between 2003 and 2021, and varied between questionnaires

RESULTS

- Among 52 sources that reported patient counts, approximately 17,355 prevalent patients per year were captured across 6 malignancies and spanning 10 countries (Table 1).
- Among the 46 sources that reported on data availability, data on patient and disease characteristics, labs, clinical outcomes, and treatment pathway were generally readily available (collected in approximately 100%, 87%, 91%, and 91% of sources, respectively) (Figure 1).
- IPI scores, FLIPI scores, and performance statuses – either ECOG scores or Karnofsky indexes – were the most well captured disease characteristic measures and were either collected or able to be derived for 92%, 88%, and 91% (respectively) of data sources with confirmed responses to outreach on availability of these variables. Charlson comorbidity indexes and CL-IPI scores were much less readily available (67% and 64%, respectively).
- Among clinical outcomes, Ann Arbor disease staging was almost always collected or able to be derived (available in 96% of data sources with a confirmed response on this variable). Number and size of lesions were only collected or able to be derived in 62% and 65% of data sources, respectively.
- Data quality and completeness was reasonable, but mixed for response measures and prognostic factors, particularly at later lines of therapy.
- Overall, genetic markers, quality of life (QoL) and healthcare resource utilisation (HCRU) measures were often not readily available (collected in only approximately 46%, 20% and 11% of sources, respectively) (Figure 1).
- Regarding operational aspects, hospitals were more willing than registries and claims databases to share patient-level data with external researchers. Most data sources reported data collection could be conducted via electronic extraction (e.g., electronic medical records, 65%) or a mix of electronic and manual methods (e.g., case report forms, 48%).
- Though data availability across the assessed European haematology-oncology data sources is sufficient to support external comparators, artificial intelligence and machine learning were identified as useful tools to help enhance real-world data. Specifically, natural language processing techniques can be used to extract unstructured data from clinical notes or pathology reports, instead of relying on more burdensome manual chart review methods.
- Given that this investigation was retrospectively conducted using information from outreach undertaken as part of routine study activities, it is important to note that the number of prevalent patients presented here is affected by a sampling bias towards indications with a greater research focus, and also towards malignancies where the findings from study activities were able to be readily compiled in the format of the research network. As such, the patient prevalence reported from this research does not comprehensively illustrate the opportunity for real-world research in malignancies like multiple myeloma that are relatively under-represented in this sample of data sources. Patient counts for haematology-oncology indications throughout Europe likely align more closely to the actual prevalence of these malignancies. Further quantitative validation of the prevalence of haematological malignancies in these data sources would be helpful to frame the landscape of available patient data in Europe.

Figure 1. Percentage of 46 assessed data sources that captured each of 7 variable categories

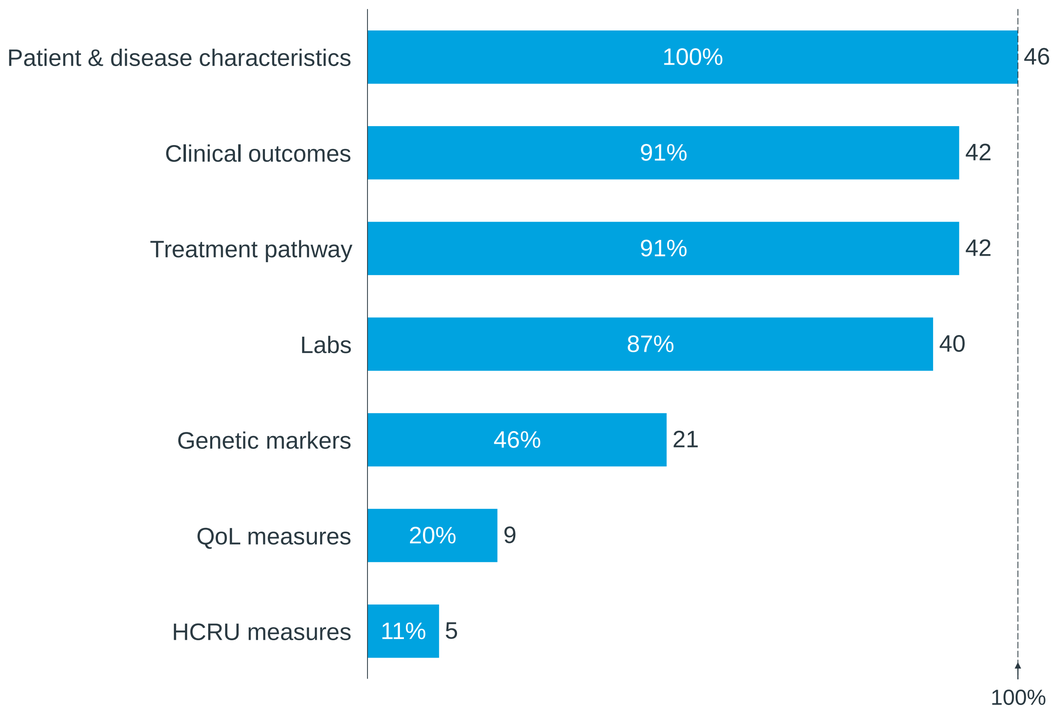

Table 1. Average prevalent patients per year captured across 6 malignancies in 10 European countries

Country (no. of data sources)**	Prevalent patients per year (no. of data sources)						Total*
	DLBCL	FL	MCL	MZL	MM	CLL	
Austria (2)	435 (2)	259 (2)	1 (1)	NA (0)	NA (0)	10 (1)	705
Czech Republic (1)	NA (0)	NA (0)	NA (0)	NA (0)	NA (0)	131 (1)	131
France (6)	1150 (6)	674 (6)	8 (1)	69 (1)	715 (3)	NA (0)	2616
Germany (9)	1319 (8)	602 (8)	78 (3)	NA (0)	1290 (4)	168 (3)	3457
Italy (7)	201 (5)	100 (5)	NA (0)	NA (0)	311 (3)	NA (0)	612
Poland (4)	205 (4)	135 (4)	NA (0)	NA (0)	NA (0)	NA (0)	340
Portugal (1)	160 (1)	78 (1)	16 (1)	NA (0)	186 (1)	49 (1)	489
Romania (1)	23 (1)	5 (1)	NA (0)	NA (0)	NA (0)	NA (0)	28
Spain (8)	303 (6)	145 (6)	21 (4)	52 (1)	388 (3)	62 (3)	971
United Kingdom (13)	2393 (13)	1136 (8)	581 (4)	20 (2)	358 (2)	3518 (3)	8006
Total*	6189	3134	705	141	3248	3938	**17355**

**Total patient numbers reported are approximate due to the normalization and scaling conducted*
***The number of data sources listed represents all data sources that reported patients within a specific country, however, not all data sources were asked to report patients for each indication listed*

NA – Patient counts were not available where outreach had not been conducted in several malignancies/countries

CONCLUSIONS

Advances are needed to facilitate real-world data collection that matches trial data more closely. To achieve this, challenges relating to variable capture, the accessibility and availability of patient-level data will need to be overcome.

Curated evidence generation networks that bring together data science and operational expertise, high quality data sources, and technological capabilities such as natural language processing could be promising tools for addressing needed improvements in data quality and operational efficiency for future haematology-oncology external comparator study execution.

ISPOR Europe 2022, 6-9 November 2022, Vienna, Austriaw

Abbreviations: CLL = Chronic lymphocytic leukaemia; DLBCL = Diffuse large B-cell lymphoma; DS = Data source; EC = External comparator; FL = Follicular lymphoma; FLIPI = Follicular lymphoma international prognostic index; HCRU = Healthcare resource utilisation; IPI = International prognostic index; MCL = Mantle cell lymphoma; MM = Multiple myeloma; MZL = Marginal cell lymphoma; QoL = Quality of life;