Received: 20 June 2024 | Accepted: 14 August 2024

DOI: 10.1111/bjh.19732

BJHaem
BRITISH JOURNAL OF HAEMATOLOGY

ORIGINAL PAPER

Haematological Malignancy – Clinical

Evaluation of coverage, generalisability and validity of the U-CAN lymphoma biobank in Sweden: A comparison with nationwide registers

Elin Forsgren[1,2] | Sara Ekberg[1,3] | Karin E. Smedby[3] | Patrick Nylund[1] | Tobias Sjöblom[1] | Max Flogegård[4] | Sara Sjöström[5] | Magnus Hultdin[6] | Karin Hallén[7] | Mats Hellström[1] | Daniel Molin[8] | Gunilla Enblad[8] | Ingrid Glimelius[1]

[1]Department of Immunology, Genetics and Pathology, Cancer Precision Medicine, Uppsala University, Uppsala, Sweden

[2]Department of Medicine, Örnsköldsvik Hospital, Örnsköldsvik, Sweden

[3]Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

[4]Department of Medicine, Falu Lasarett, Falun, Sweden

[5]Oncology Section, Department of Diagnostics and Intervention, Norrland University Hospital, Umeå, Sweden

[6]Pathology Section, Department of Medical Biosciences, Umeå University, Umeå, Sweden

[7]Department of Oncology, Karlstad Hospital, Karlstad, Sweden

[8]Department of Immunology, Genetics and Pathology, Cancer Immunotherapy, Uppsala University, Uppsala, Sweden

Correspondence
Elin Forsgren, Department of Immunology, Genetics and Pathology, Cancer Precision Medicine, Uppsala University, Uppsala, Sweden.
Email: elin.forsgren@igp.uu.se

Funding information
Swedish Research Council, Grant/Award Number: 2022-00801; Swedish Cancer Society, Grant/Award Number: CAN 222167Pj; The Sjöberg Foundation, Grant/Award Number: 2023-01-03:3

Summary

Validation of biobanks and large cancer cohorts is essential in ensuring high-quality research results. We examined the coverage, generalisability and validity of the lymphoma collection of the Uppsala-Umeå Comprehensive Cancer Consortium (U-CAN) biobank in Sweden, one of the largest cancer biobanks in Europe. Up until 2022, 889 lymphoma patients in U-CAN Uppsala had available samples, and 329 in U-CAN Umeå. Patients diagnosed in the U-CAN Uppsala area 2011–2021 ($n = 843$) were linked to the nationwide Swedish Lymphoma Register, and a subset diagnosed before 2019 ($n = 727$) to population-based registers. The coverage was 39% of all lymphoma patients between 2011 and 2019 diagnosed in the U-CAN Uppsala area, with a pandemic decline to 10% during 2020–2021. The patients included had superior overall survival (hazard ratio = 0.70 [95% confidence interval, CI: 0.60–0.82]) than all lymphoma patients in Sweden. They had better performance status, were younger (odds ratio [OR] = 0.21 [95% CI: 0.13–0.34]) and had less comorbidities (OR = 0.66 [95% CI: 0.56–0.78]). However, cause-specific survival and stage distribution were similar. The questionnaire data captured less comorbidities compared to the national registers. Evaluations of biobanks are important, as even population-based biobanks such as U-CAN select younger patients with higher socioeconomical status and better performance status. However, the similar cause-specific survival as in the registries suggests U-CANs usefulness for prognostic biomarker studies.

KEYWORDS
biobank, comorbidity, lymphoma, mortality

INTRODUCTION

Validating large cancer cohorts and biobanks is crucial in ensuring high-quality research and generalisability to a broader population.[1] Biobanks may have a non-representative population, especially regarding age, access to treatment and socioeconomic aspects, which has, for example, been described in the United Kingdom (UK) biobank.[2] It is important to assess the external validity of research cohorts to evaluate the generalisability of results obtained for the

cohorts. However, the ability to conduct such studies is often restricted by the absence of nationwide register-based data, which is available in the Nordic countries.

Uppsala-Umeå Comprehensive Cancer Consortium (U-CAN), initiated in 2010, is an ongoing prospective longitudinal collection of biospecimens accompanied by clinical information from adult cancer patients, and it is one of the largest cancer biobanks in Europe.[3] Biobanks allow for the collection of population-based materials, and the strict inclusion and exclusion criteria often used in clinical trials are not needed. However, included patients may still not be entirely representative of the underlying patient population. To date, samples from the U-CAN biobank have been utilised in approximately 150 research studies. Novel biomarkers have been identified for prognostication, treatment prediction and immunotherapy response in diffuse large B-cell lymphoma,[4] and plasma-based immune-proteasome profiling and serum-based biomarker discovery analyses have been instrumental in distinguishing between malignant cell transformation and normal control tissue in Hodgkin and diffuse large B-cell lymphoma.[5–8]

As biobanks and research data often rely on self-reported information on comorbidities and socioeconomic factors,[3,9] misclassification may occur. While nationwide register-based data are not susceptible to this type of misclassification, they do present other limitations, such as lack of detailed information and life-style factors. Utilising the questionnaire data collected at inclusion in U-CAN Uppsala and the corresponding International Statistical Classification of Diseases and Related Health Problem-codes (ICD-codes) recorded in the Swedish National Patient Register (NPR), offers a unique opportunity to compare different approaches to collecting comorbidity data.[10] In addition, assessments of the U-CAN biobank's coverage can be conducted using data from the nationwide Swedish Lymphoma Register (SLR). Finally, sociodemographic variables reported in questionnaires can be cross-referenced with data from the longitudinal integrated database for health insurance and labour market studies (LISA) in Sweden.[11]

The aim of this study was to assess the coverage and generalisability of the U-CAN lymphoma biobank within the Uppsala catchment area and describe the total number of patients included and available samples. Additionally, we aimed to evaluate the validity of self-reported data on education level, comorbidity and marital status by comparing results to population-based registers in Sweden.

METHODS

U-CAN biobank

This prospective biobank was established based on a long-standing collaboration between Uppsala and Umeå University, Uppsala Biobank and Biobanken Norr, and comprises several cancer diagnoses. In the lymphoma cohort, patients are first diagnosed with a biopsy confirming lymphoma. Next, patients are invited to participate during their clinical visit prior to deciding on management of the patient. Upon signing informed consent, blood and tissue samples are collected at diagnosis and subsequently at regular intervals (depending on the cancer/lymphoma type) (Figure 1). In the event of a relapse, new blood and tissue samples are obtained. Patients diagnosed with lymphoma have been recruited from 2010, with the only exclusion criteria being under 18 years of age. Upon inclusion, patients are asked to complete a questionnaire regarding heredity and lifestyle factors, including educational level, marital/partner status, self-reported performance status, and any additional illnesses or comorbidities (Data S2).

The Swedish Lymphoma Register

The SLR was established in 2000 and encompasses approximately 95% of lymphoma cases in Sweden.[12,13] Patients with chronic lymphocytic leukaemia (CLL) are not included. In addition to recorded diagnosis, treatment details, and relapses, the register also collects supplementary clinical information such as Ann Arbor stage and laboratory parameters at diagnosis.

Study population

Since this study required data linkages, further results were based on the U-CAN Uppsala cohort and did not include the U-CAN Umeå cohort. To establish the study cohort, 1090 patients diagnosed with lymphoma or CLL were enrolled from the U-CAN hospitals in the Uppsala catchment area spanning from 2010 until data extraction in December 2022. To enable comparison to SLR, patients with CLL ($n = 154$) and those included at the time of relapse ($n = 44$) were excluded from this evaluation. In addition, three patients lacked information on inclusion dates or diagnostic dates. Following these exclusions, a refined cohort of 889 U-CAN patients was established (Figure 2), from which we present the available samples. Further analyses required linkages through registries, which had been performed up until 2021.

Coverage of U-CAN in Uppsala catchment area 2011–2021

Data from the SLR was available until 2021. To study the coverage of U-CAN in the Uppsala catchment area, we compared the patients included in U-CAN at the four U-CAN hospitals, to all lymphoma patients diagnosed and recorded in the SLR at these hospitals during the specific recruitment period: Uppsala (years 2011–2021), Falun (1st of May 2012 to 31st of December 2021), Karlstad (1st of October 2014 to 31st of December 2021) and Gävle (years 2016–2021). We excluded patients diagnosed under 2010, as the U-CAN biobank was established during that year. Between 2011 and

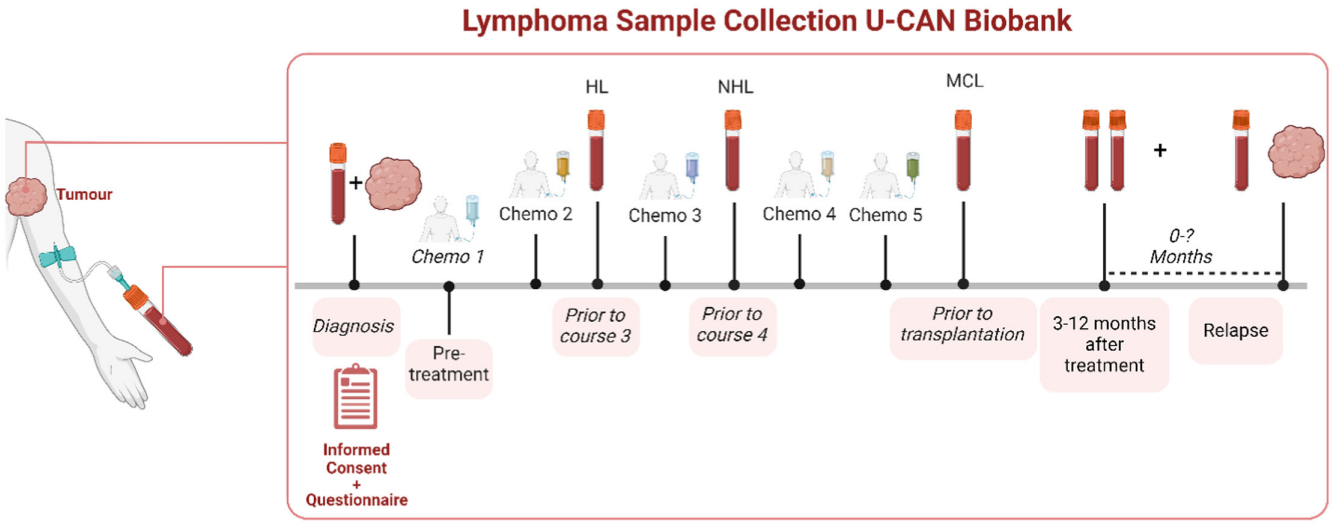

FIGURE 1 Procedures for sampling of data, blood and tissue are standardised across different diagnoses in the U-CAN biobank. Blood and tissue samples and a questionnaire encompassing heredity and lifestyle are collected for lymphoma patients at diagnosis. Depending on indolent or aggressive lymphoma type[a], blood samples are continuously collected during the disease course, as outlined above. Samples collected in the U-CAN Biobank include one whole blood sample, two plasma samples and one serum sample. In addition, all patients included in the U-CAN biobank have their original as well as potential relapsed tumour biopsy stored and registered at the pathology department, enabling research within the U-CAN framework. [a]For HL samples are taken prior to course three to match the planned fluorodeoxyglucose positron emission tomography scan. Samples prior to transplantation are preferably taken for patients with MCL and T-cell lymphoma. chemo, chemoimmunotherapy; HL, Hodgkin lymphoma; MCL, mantle cell lymphoma; NHL, non-Hodgkin lymphoma; U-CAN, Uppsala-Umeå Comprehensive Cancer Consortium.

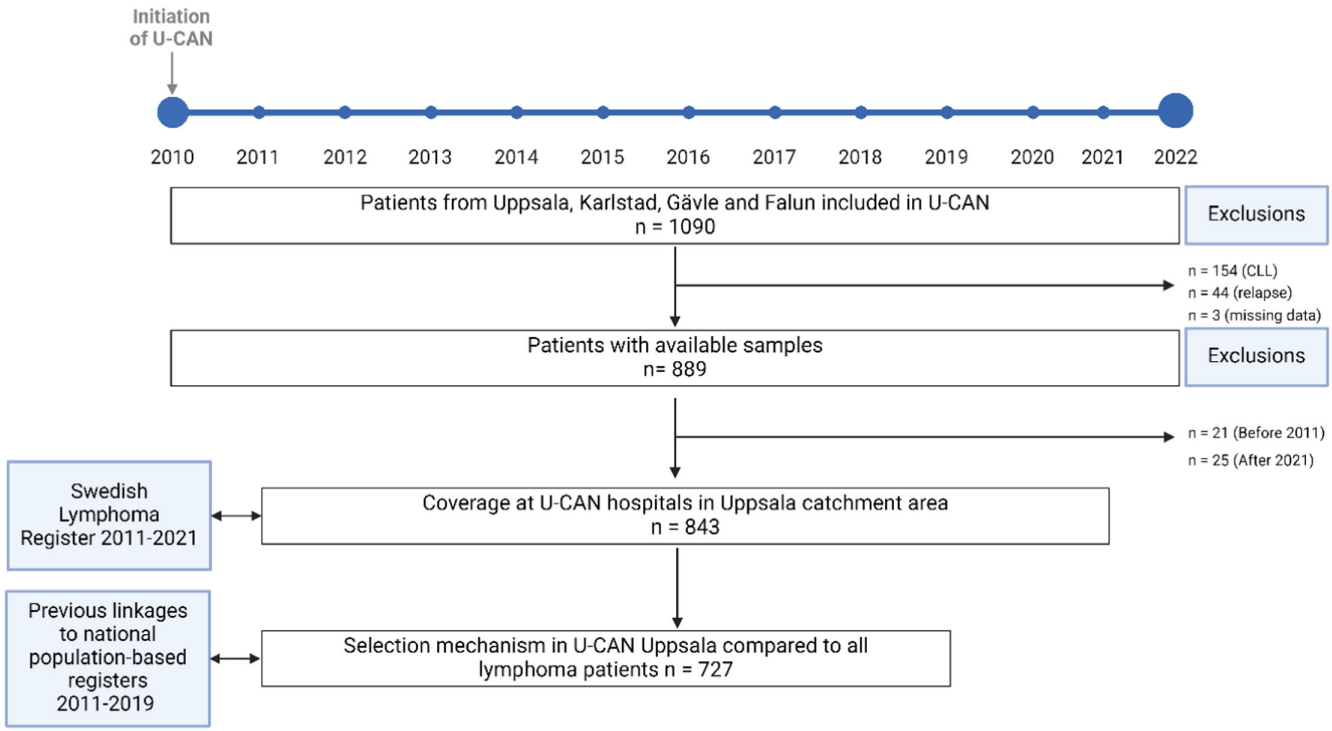

FIGURE 2 Description of the study cohort of lymphoma patients diagnosed and included in U-CAN Uppsala and the comparison group from the Swedish Lymphoma Register (SLR). For a relevant comparison of patients in the U-CAN biobank to SLR, the following exclusions were made. Patients with chronic lymphocytic leukaemia (CLL) were excluded, as they are not included in the SLR, as well as patients included in U-CAN outside the inclusion window and patients included at relapse. Patients diagnosed up to 2019 had previously been linked to population-based registers such as Swedish National Patient Register, Longitudinal Integrated Database for Health Insurance and Labour Market Studies and Cause of Death Register. U-CAN, Uppsala-Umeå Comprehensive Cancer Consortium.

2019, the coverage analysis was performed on an individual level, through linkages to registries, and between 2020 and 2021, the coverage analysis was performed on group level, as we missed the personal linkages due to the waiting period for the data processing time by the Swedish authorities. The comparison involved examining the number of patients stratified by age, sex, year of diagnosis and lymphoma subtype for the corresponding years of each hospital's inclusion time frame.

Linkage to other national population-based registers

A subset of the U-CAN data ($n = 727$, 86%), corresponding to patients diagnosed until 2019, could be individually linked to other population-based registers besides the SLR using the Swedish personal identification number.[14] These registers include the Swedish NPR, including both inpatient hospital care and outpatient non-primary care, making it possible to calculate the Charlson Comorbidity Index (CCI). Additionally, information on educational level (categorised as ≤9 years, 10–12, or ≥13 years of schooling) and partnership (categorised as married, never married, divorced or widowed) was retrieved from the national database LISA.[11] Educational level was used as a proxy for socioeconomic status.[15] As it is forbidden to gather information on ethnicity according to the Swedish Personal Data Act (Personsuppgiftslagen 1998:204), no such linkages could be performed. Date of death and causes of deaths were obtained from the Cause of Death Register.

Selection mechanism

To understand the characteristics of the U-CAN lymphoma cohort in comparison to the broader lymphoma population, we compared patients included in U-CAN up until 2019 ($n = 729$) with those in the SLR in all of Sweden. Logistic regression models were utilised to estimate odds ratios (OR) with 95% confidence intervals (CIs) to evaluate the impact of various clinical and socioeconomic characteristics on participation in U-CAN versus patients in SLR in all of Sweden. Furthermore, we compared all-cause and cause-specific (i.e. lymphoma-specific) survival among U-CAN participants to that of the broader SLR population using the Kaplan–Meier method. Patients were followed from the date of diagnosis until death (due to any cause or due to lymphoma) or until the 31st of December 2019, whichever occurred first. Cox regression models were employed to estimate hazard ratios (HRs) with 95% CI as an assessment of the influence of selection on survival outcomes. To adjust for potential confounding factors, HRs were also adjusted for age, sex, CCI and marital/partnership status.

Validation of self-reported medical information

The patient questionnaire solicited information on both historical and existing health conditions including diabetes mellitus, hypertension, hypothyroidism, biliary tract diseases, liver diseases, myocardial infarction, angina pectoris, heart failure, ischaemic and haemorrhagic stroke, lung diseases, intestinal diseases, kidney diseases and prior occurrence of cancer. To evaluate the reliability of self-reported medical information gathered through the questionnaire, we cross-referenced patients' responses with their records of pre-diagnostic outpatient and inpatient care as documented in the NPR, utilising corresponding ICD-codes (Table S1). To quantitatively assess the sensitivity of the U-CAN data, we computed the ratio of patients who reported a specific medical condition in the questionnaire to the total number of U-CAN patients with corresponding matches in the NPR records for the same condition. Concurrently, specificity was determined by calculating the ratio of the number of patients who *did not* self-report a specific condition to the total number of patients without corresponding matches in the NPR records, as indicated by the relevant ICD codes. Similarly, self-reported educational level and civil status were compared to information from the LISA database.

Statistical analysis

All statistical analyses were performed using Stata v.17 (StataCorp. 2019. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC).

RESULTS

Available tissue samples in U-CAN lymphoma up until 2022

At the time of data extraction, 889 lymphoma patients diagnosed between 2010 and 2022 were available in the Uppsala cohort, with 100% having a pathology specimen, 99% having at least one blood sampling time point (whole blood, serum and plasma, $n = 883$), and 67% of patients included had multiple sampling time points available ($n = 592$) (Table S2). A parallel gathering of biospecimens in U-CAN Umeå included 329 lymphoma patients (Table S2).

The coverage of enrolment in U-CAN Uppsala between 2011 and 2021

The coverage analysis was performed on patients up until 2021, since the SLR data were available until then. Between 2011 and 2019, 39% of lymphoma patients diagnosed at the four U-CAN hospitals in the Uppsala catchment area were

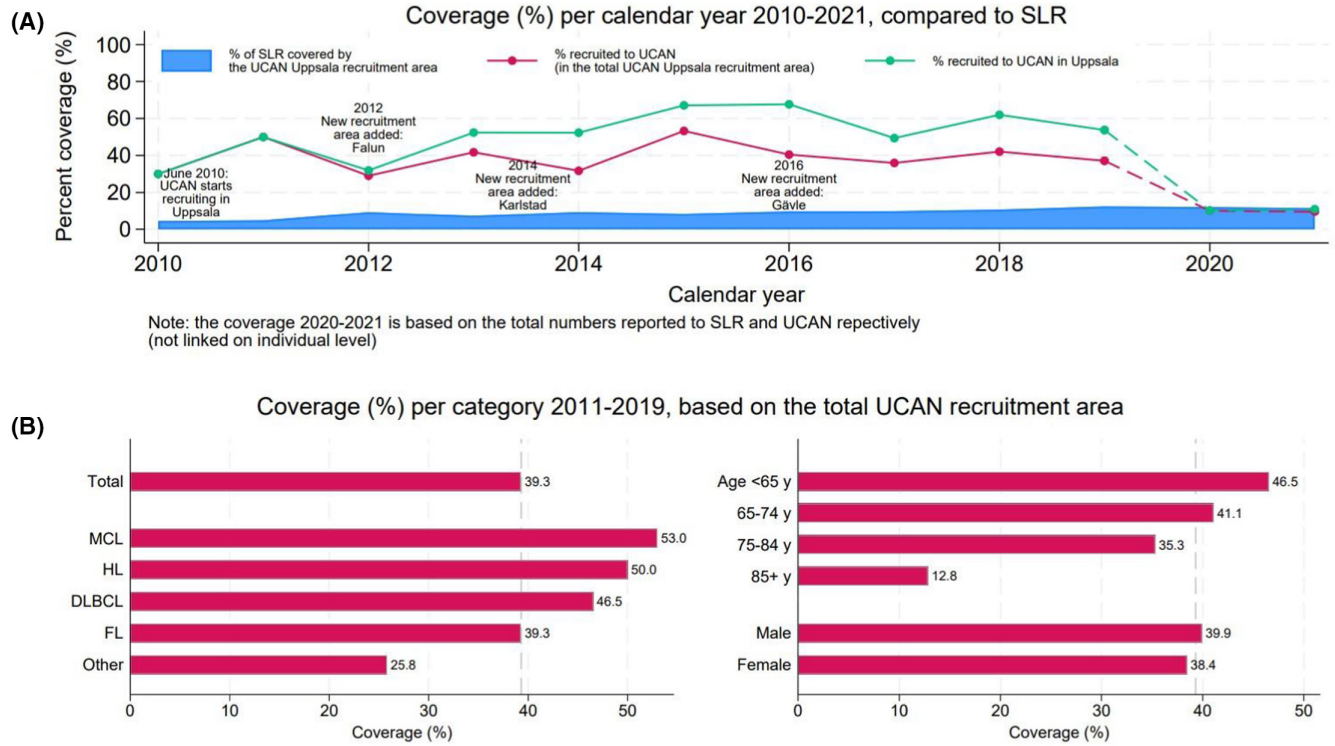

FIGURE 3 (A) The coverage of U-CAN lymphoma in the Uppsala catchment area between 2011 and 2021. (B) Percentage of enrolment in U-CAN between 2011 and 2019 in the Uppsala catchment area compared to the Swedish Lymphoma Register, stratified by calendar year diagnosis and age. (A) The number of patients included in U-CAN ($n = 843$) at the four U-CAN hospitals was compared to all patients diagnosed in the Swedish Lymphoma Register at these hospitals during the specific recruitment period: Uppsala (years 2011–2021), Falun (1st of May 2012 to 31st of December 2021), Karlstad (1st of October 2014 to 31st of December 2021) and Gävle (years 2016–2021). In patients diagnosed in 2020–2011, individual data linkages were not performed. The analysis was not performed on patients diagnosed in 2010 as the biobank was established during that year. (B) The number of patients included in U-CAN ($n = 796$) at the four U-CAN hospitals were compared to all patients diagnosed in the Swedish Lymphoma Register at these hospitals during the specific recruitment period, stratified by diagnosis, age and gender, during 2011–2019. U-CAN, Uppsala-Umeå Comprehensive Cancer Consortium.

included in U-CAN (Figure 3). However, coverage markedly declined to 10% during 2020 and 2021 (Figure 3A). Notably, the coverage within U-CAN was higher for patients diagnosed at the Uppsala university hospital compared to the other hospitals, among younger patients and those diagnosed with mantle cell lymphoma or Hodgkin lymphoma compared to other lymphoma subtypes. Coverage did not vary markedly by sex (Figure 3B).

Selection mechanisms in U-CAN compared to lymphoma patients in Sweden during 2011–2019

The selection analysis was performed on patients up until 2019, since the personal linkages to registries were performed until then. Compared to all lymphoma patients in Sweden diagnosed between 2011 and 2019, patients included in U-CAN during the same period were younger (e.g. age 85+ years vs. ≤65 years: OR = 0.21 [95% CI 0.13–0.34]), had higher education level (e.g. ≥13 years of education vs. ≤9 years: OR = 1.41 [95% CI 1.14–1.68]), fewer comorbidities at diagnosis (CCI 2+ vs. 0: OR = 0.66 [95% CI 0.56–0.78]) or were more likely to have a partner (partner vs. no partner: OR = 1.21 [95% CI 1.04–1.41]) (Table 1). Additionally, patients included in U-CAN exhibited better performance status (ECOG 2–4 vs. 0–1: OR = 0.55 [95% CI 0.43–0.71]) and were more likely to have normal serum lactate dehydrogenase (S-LDH) levels at diagnosis (elevated vs. normal S-LDH: OR = 0.68 [95% CI 0.58–0.81]) (Table 1). Overall, there was no difference in Ann Arbor stages between patients included in U-CAN and in the SLR. However, when considering indolent and aggressive lymphoma subtypes separately, the indolent lymphomas were more likely to be included in U-CAN if diagnosed at more advanced stages (stage II or III, e.g. Ann Arbor stage III vs. I: OR = 1.87 [95% CI 1.10–3.18]) (Tables S3 and S4).

Survival outcomes between the U-CAN cohort and patients in the SLR throughout Sweden

Patients included in U-CAN exhibited a better all cause survival rate compared to patients in the SLR (HR = 0.70 [95% CI: 0.60–0.81]), but the difference became insignificant after adjustment (age, sex, CCI, education level and marital status) (Figure 4). Furthermore, the cause-specific survival was similar between the U-CAN cohort and the

TABLE 1 Characteristics of patients included in U-CAN in the Uppsala catchment area 2011–2019, compared to characteristics of lymphoma patients in the whole of Sweden registered in the Swedish Lymphoma Register (SLR) during the same calendar years.

	SLR		Included in U-CAN		
Patient characteristics	***N***	**%**	***N***	**%**	**ORR (95% CI)**
Sex					
Male	11 261	57.1	422	58.0	1.00 (Reference)
Female	8454	42.9	305	42.0	0.96 (0.83, 1.12)
Age at diagnosis					
<65	7036	35.7	326	44.8	1.00 (Reference)
65–74	5771	29.3	227	31.2	0.84 (0.71, 1.00)
75–84	5056	25.6	155	21.3	**0.65** (**0.54**, **0.79**)
85+	1852	9.4	19	2.6	**0.21** (**0.13**, **0.34**)
Highest achieved education level					
≤9	5974	30.3	188	25.9	1.00 (Reference)
10–12	7997	40.6	307	42.2	**1.23** (**1.02**, **1.48**)
≥13	5280	26.8	231	31.8	**1.41** (**1.16**, **1.71**)
Missing	464	3.5	22	3.0	
Partner status at diagnosis					
No	8448	42.9	282	38.8	1.00 (Reference)
Yes	10 569	53.6	423	58.2	**1.21** (**1.04**, **1.41**)
Missing	698	3.5	22	3.0	
Type of hospital					
University hospital	7386	48.7	418	58.0	1.00 (Reference)
Non-university hospital	6185	38.5	296	41.1	0.89 (0.77, 1.04)
Missing	2058	12.8	7	1.0	
CCI, categorised					
0	9066	46.2	390	53.8	1.00 (Reference)
1	2633	13.4	106	14.5	0.92 (0.74, 1.15)
2+	7931	40.4	230	31.7	**0.66** (**0.56**, **0.78**)
Stage (Ann Arbor)					
I	3536	17.9	151	20.8	1.00 (Reference)
II	2916	14.8	133	18.3	1.07 (0.84, 1.36)
III	2567	13.0	115	15.8	1.05 (0.82, 1.35)
IV	7314	37.1	285	39.2	0.91 (0.74, 1.11)
Missing	3383	17.2	43	5.9	
Elevated S-LD					
No	8514	43.2	407	56.0	1.00 (Reference)
Yes	6565	33.3	217	29.8	**0.68** (**0.58**, **0.81**)
Missing	4636	23.5	103	14.2	
ECOG					
0–1	14 557	73.8	650	89.4	1.00 (Reference)
2–4	2665	13.5	67	9.2	**0.55** (**0.43**, **0.71**)
Missing	2493	12.6	10	1.4	
B symptoms					
No	11 443	58.0	507	69.7	1.00 (Reference)
Yes	5093	25.8	201	27.6	0.89 (0.75, 1.05)
Missing	3179	16.1	19	2.6	

(Continues)

TABLE 1 (Continued)

Patient characteristics	SLR		Included in U-CAN		
	N	%	*N*	%	ORR (95% CI)
Bulky disease					
No	14 469	73.4	591	81.3	1.00 (Reference)
Yes	2383	12.1	120	16.5	**1.25 (1.02, 1.52)**
Missing	2863	14.5	16	2.2	

Note: Odds ratios and 95% confidence intervals are presented for the likelihood of inclusion versus non-inclusion in U-CAN. Significant values are in bold.

Abbreviations: CCI, Charlson Comorbidity Index; ECOG, Eastern Cooperative Oncology Group Performance status; S-LD, serum lactate dehydrogenase level; U-CAN, Uppsala-Umeå Comprehensive Cancer Consortium.

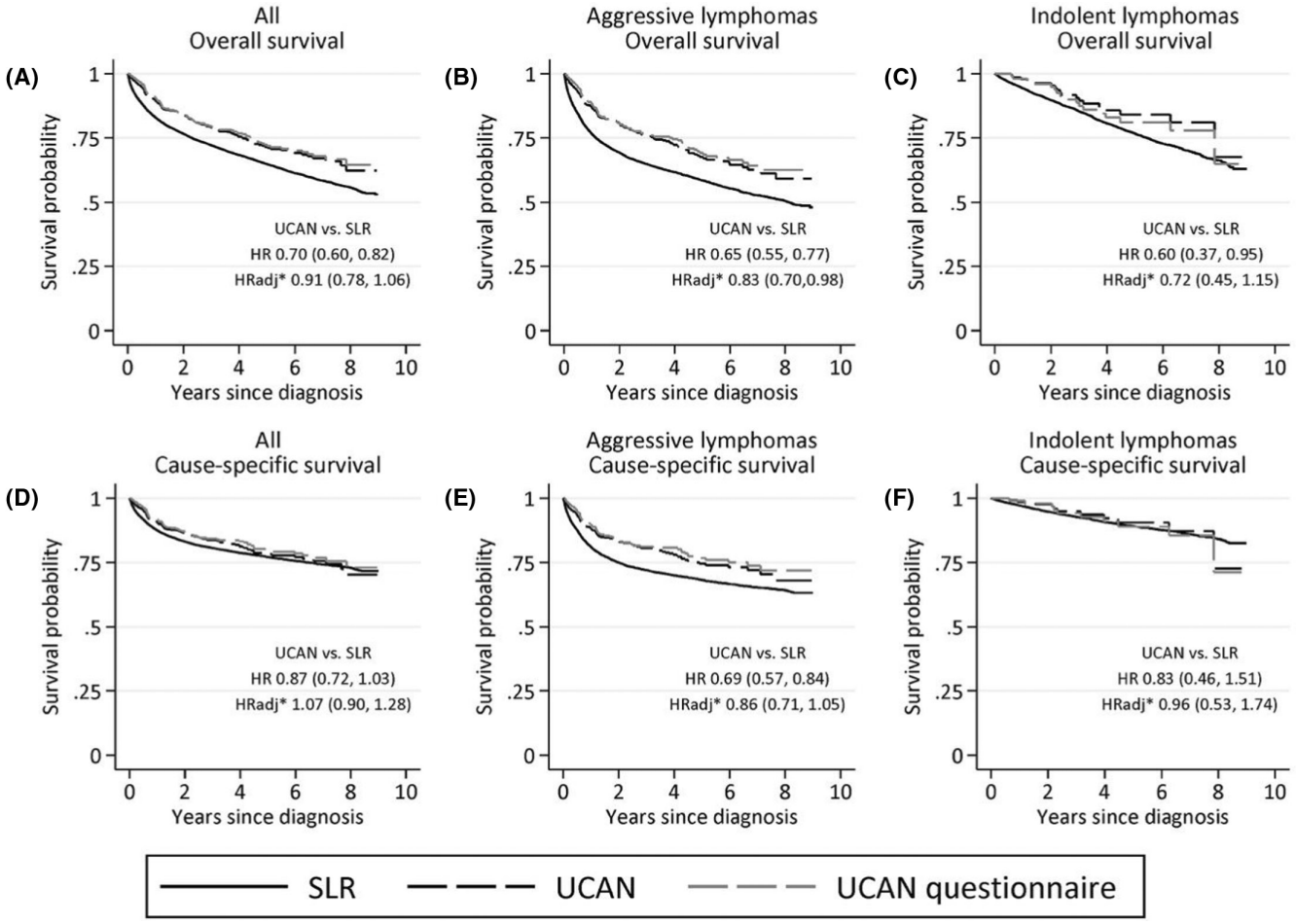

FIGURE 4 Survival probabilities and hazard ratios (HRs) (unadjusted (HR) and adjusted (HRadj*) for age, sex, Charlson Comorbidity Index, education level and marital status) of overall survival. (A–C), and cause-specific survival (D–F), comparing lymphoma patients in U-CAN (with or without questionnaire) to all other patients in the Swedish Lymphoma Register 2011–2019. U-CAN, Uppsala-Umeå Comprehensive Cancer Consortium.

SLR patients. However, patients with aggressive lymphomas had a better cause-specific survival than the aggressive lymphoma patients included in the SLR (HR = 0.69 [95% CI: 0.57–0.84]) when not adjusted. The indolent lymphoma patients in U-CAN demonstrated a survival rate similar to the indolent patients included in the SLR (HR = 0.83 [95% CI: 0.46–1.51]).

When comparing the 5-year probability of death, there was no difference in death due to lymphoma (U-CAN 21.1% vs. SLR 21.6%) (Figure S2). The U-CAN patients were less likely to die of cardiovascular disease and other malignancies than patients included in the SLR.

Comparison between questionnaire data in U-CAN and national registers

Out of the 727 U-CAN patients diagnosed until 2019 and linked to other national registers, 526 (72%) responded to the questionnaire. The response rates did not vary markedly

by sex, educational level, lymphoma subtype or Ann Arbor stage (Figure S1). However, the oldest patients and those with the worst performance status had a lower response rate to the questionnaire. Importantly, there was no difference in the all-cause or cause-specific survival between U-CAN patients who answered the questionnaire and patients who did not (Figure 4).

Comparing the self-reported diseases in the U-CAN questionnaire to the diseases registered as ICD-codes in the NPR, patients did not report illnesses to the same extent as recorded in the national registers, especially mental disease and kidney disease. However, diagnoses such as diabetes and hypertension were more likely to be reported (Figure 5). When comparing the absence of self-reported disease in U-CAN with the absence of records in national registers, very few patients reported illnesses that were not recorded in the national registers (Figure 5).

DISCUSSION

In this retrospective cohort study conducted in Sweden, the lymphoma cohort within the U-CAN Uppsala catchment area demonstrated a coverage of 39% between 2011 and 2019, with a decline during 2020–2021. This coverage rate may be lower than anticipated, given U-CAN's goal to have a population-based inclusion rate. Achieving complete coverage may be unfeasible under conditions demanding expedited clinical management, as patients must be adequately informed and able to sign an informed consent to participate in the biobank process. Notably, a coverage of 53% in mantle cell lymphoma and 50% in Hodgkin lymphoma patients can still be considered relatively high. In addition, we observed a distinct impact of the pandemic years on U-CAN inclusion rates, with coverage dropping to 10% during 2020 and 2021. This decline aligns with previous observations resulting in a reduction in newly diagnosed patients in clinical trials during the pandemic.[16] However, in a previous Swedish study, only a slight decline of lymphoma diagnoses was observed nationally during the first few months of the pandemic.[17]

Our study highlights the challenge of achieving completely non-selected biobank sampling, despite ambitious goals. The U-CAN lymphoma cohort has lower representation of elderly patients and patients from lower socioeconomic groups. Previous research on the UK biobank has revealed a tendency towards a 'healthy volunteer effect', where participants exhibit lower mortality rates compared to the general population.[18] This has also been displayed in a large, prospective lymphoma cohort (the LEO cohort) in the United States, where the research cohort was underrepresented for patients who were aged 80 years and above.[19] Similar to our study, they observed better survival in aggressive lymphomas in the research cohort compared to population-based data. In other large biobanks such as the Victorian Cancer

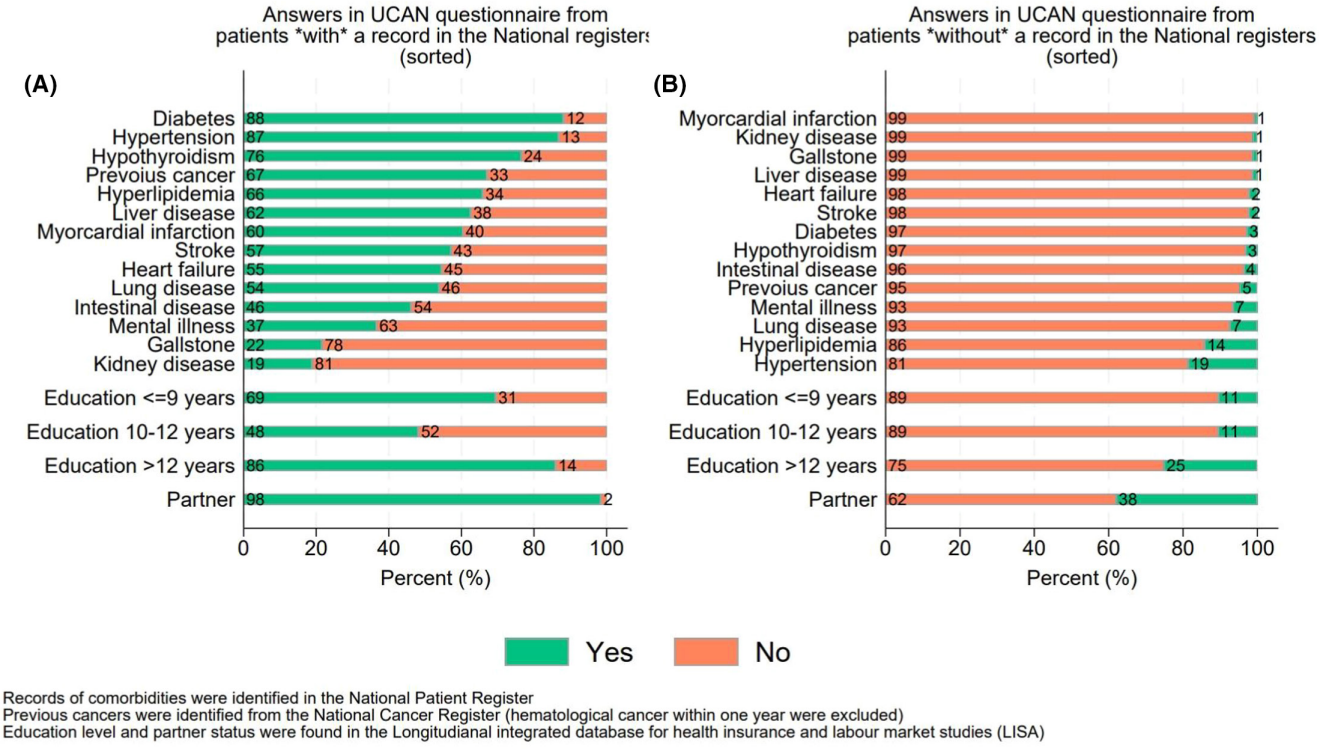

FIGURE 5 Comparison of the self-reported diseases for patients who answered the U-CAN questionnaire ($n = 526$) and their records in the National Patient Register (as International Classification of Disease-codes). (A) The per cent (%) of patients with a reported disease versus non-reported disease in U-CAN if there was a record in the national registers. (B) The per cent (%) of patients with a reported disease versus non-reported disease in U-CAN if there was no record in the national registers. U-CAN, Uppsala-Umeå Comprehensive Cancer Consortium.

Biobank in Australia and the Wales Cancer Bank, assessments of coverage and potential selection bias of participants compared with the general lymphoma population have not been conducted.[20,21]

Our study underscores the importance of evaluating the performance of biobanks utilised for cancer research, as disparities in comorbidities, genetic variation and underrepresented patient groups may negatively affect the generalisability of new findings in selected cohorts. However, a thoroughly evaluated research cohort can provide guidance on potential pitfalls. We identified certain selection mechanisms regarding socioeconomic factors and comorbidities but the similar cause-specific overall survival and the same 5-year probability of death in lymphoma implying a comparable disease course between patients included in U-CAN and those who were not. Until now, the primary focus of studies utilising the U-CAN lymphoma cohort has been directed towards exploration of new biomarkers. There is currently no evidence that lower socioeconomic status and comorbidities significantly influence tumour biology or disease progression. However, these factors can influence prognosis in real-world settings, as they may affect treatment tolerance and overall survival. Therefore, validation of new biomarker studies in these patient groups would be of interest (Figure S3).

As the treatment landscape in haematological malignancies shifts towards personalised and non-chemotherapeutic approaches, the linkage between biological and epidemiological data becomes increasingly relevant. In this regard, U-CAN has a unique potential as a biobank with its linkages to national registers. Our findings indicate that patients tend to report fewer diagnoses than are registered in the NPR, particularly concerning mental health and kidney diseases. Even severe conditions such as heart failure, myocardial infarction and previous cancer exhibited low sensitivity. Consequently, our results suggest that questionnaire data should be interpreted cautiously, and national registers may serve as more reliable sources of comorbidity and socioeconomic status information for certain diseases.

The strengths of our study lie in its inclusion of a well-documented cohort and the ability to link clinical data to national registries. However, some limitations should be acknowledged. Firstly, our study only covers the Uppsala area of U-CAN (the hospitals in Uppsala, Falun, Karlstad and Gävle). Secondly, this restricted geographic scope may limit the generalisability of our findings as we compared the U-CAN Uppsala lymphoma cohort with all lymphoma patients in Sweden. However, it is important to note that the U-CAN Uppsala area may not fully reflect the socioeconomic diversity of the entire Swedish population. This discrepancy in socioeconomic standards could contribute to the selection mechanisms identified in our study.

In conclusion, the U-CAN Uppsala biobank demonstrated a coverage of 39% among all lymphoma patients in the Uppsala catchment area during 2011–2019, with 99% of the patients with available samples. There was no difference in cause-specific survival between U-CAN and the broader lymphoma population, suggesting its usefulness for the general lymphoma population. However, the finding that biobank patients are younger and with less comorbidity, as well as report themselves with less comorbidity compared to national registries, highlights the importance of using a combination of data sources to validate and enhance the accuracy of research findings in biobank studies.

AUTHOR CONTRIBUTIONS

Ingrid Glimelius, Sara Ekberg and Elin Forsgren designed this study; Sara Ekberg did statistical analysis; Elin Forsgren and Ingrid Glimelius wrote the manuscript, Ingrid Glimelius, Mats Hellström, Daniel Molin, Gunilla Enblad, Max Flogegård and Sara Sjöström collected lymphoma patients. Tobias Sjöblom is responsible for U-CAN in Uppsala; Magnus Hultdin is responsible for U-CAN in Umeå. Patrick Nylund contributed with illustration and writing of the manuscript. Karin E. Smedby is responsible for data linkages and the Swedish lymphoma register. All authors have read and approved the manuscript.

ACKNOWLEDGEMENTS

The authors declare financial support was received for the research, authorship and/or publication of this article. The project was supported by grants from the Swedish Research Council (Dnr: 2022-00801), the Swedish Cancer Society (CAN 222167Pj) and The Sjöberg Foundation (2023-01-03:3). We thank Lars Skagerlind and Kristina Lundqvist, University Hospital, Umeå, for support with the U-CAN Umeå data.

CONFLICT OF INTEREST STATEMENT

Ingrid Glimelius has received support to the department unrelated to this project from Jansen Cilag, Takeda and Kite Gilead. Karin E. Smedby received honoraria from Incyte, Abbvie and Celgene, and research support from Janssen Cilag. Tobias Sjöblom is co-founder, board member and shareholder of Oncodia AB. The others report no conflict of interest.

DATA AVAILABILITY STATEMENT

Due to patient consent and confidentiality, the dataset presented herein can only be made available upon request to the corresponding author.

ETHICS STATEMENT

The U-CAN sample and data collection operate under ethical approval Dnr: 2010/198 (Uppsala Ethical Review Committee). Ethical approval for this study includes gathering information from the U-CAN cohort for research purposes (Dnr: 2013/059 and 2014-233) and for the Lymphoma base linkages (Dnr 2019/4:1, with amendment Dnr: 2021-06739), according to the Declaration of Helsinki.

PATIENT CONSENT STATEMENT

Patients signed informed consent at inclusion in the U-CAN biobank.

ORCID

Elin Forsgren https://orcid.org/0009-0006-0837-7410
Patrick Nylund https://orcid.org/0000-0002-1274-4010
Ingrid Glimelius https://orcid.org/0000-0001-6158-3041

TWITTER

Ingrid Glimelius glimelius

REFERENCES

1. Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. Target validity and the hierarchy of study designs. Am J Epidemiol. 2019;188(2):438–43.
2. Keyes KM, Westreich D. UK biobank, big data, and the consequences of non-representativeness. Lancet. 2019;393(10178):1297.
3. Glimelius B, Melin B, Enblad G, Alafuzoff I, Beskow A, Ahlström H, et al. U-CAN: a prospective longitudinal collection of biomaterials and clinical information from adult cancer patients in Sweden. Acta Oncol. 2018;57(2):187–94.
4. Abdulla M, Alexsson A, Sundström C, Ladenvall C, Mansouri L, Lindskog C, et al. PD-L1 and IDO1 are potential targets for treatment in patients with primary diffuse large B-cell lymphoma of the CNS. Acta Oncol. 2021;60(4):531–8.
5. Gholiha AR, Hollander P, Löf L, Larsson A, Hashemi J, Ulfstedt JM, et al. Immune-proteome profiling in classical Hodgkin lymphoma tumor diagnostic tissue. Cancer. 2021;14(1):9.
6. Mattsson Ulfstedt J, Venge P, Holmgren S, Enblad G, Eriksson S, Molin D. Serum concentrations of thymidine kinase 1 measured using a novel antibody-based assay in patients with Hodgkin lymphoma. Ups J Med Sci. 2021;126.
7. Mörth C, Sabaa AA, Freyhult E, Christersson C, Hashemi J, Hashemi N, et al. Plasma proteome profiling of cardiotoxicity in patients with diffuse large B-cell lymphoma. Cardiooncology. 2021;7(1):6.
8. Menter T, Tzankov A, Zucca E, Kimby E, Hultdin M, Sundström C, et al. Prognostic implications of the microenvironment for follicular lymphoma under immunomodulation therapy. Br J Haematol. 2020;189(4):707–17.
9. Pukkala E. Biobanks and registers in epidemiologic research on cancer. In: Dillner J, editor. Methods in biobanking [Internet]. Totowa: Humana Press; 2011. p. 127–64. https://doi.org/10.1007/978-1-59745-423-0_5
10. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011;11(1):450.
11. Ludvigsson JF, Svedberg P, Olén O, Bruze G, Neovius M. The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. Eur J Epidemiol. 2019;34(4):423–37.
12. Nationellt Kvalitetsregister Lymfom. RCC [Internet]. [cited 2022 Nov 22]. Available from: https://cancercentrum.se/samverkan/cancerdiagnoser/blod-lymfom-myelom/lymfom-lymfkortelcancer/kvalitetsregister
13. Ekström Smedby K, Eloranta S, Wästerlid T, Falini V, Jerlström U, Ellin F, et al. The National Swedish Lymphoma Register – a systematic validation of data quality. Acta Oncol. 2024;63:563–72.
14. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. Eur J Epidemiol. 2009;24(11):659–67.
15. Pálmarsdóttir R, Kiesbye Øvlisen A, Severinsen MT, Glimelius I, Smedby KE, El-Galaly T. Socioeconomic impact of Hodgkin lymphoma in adult patients: a systematic literature review. Leuk Lymphoma. 2019;60(13):3116–31.
16. Englum BR, Prasad NK, Lake RE, Mayorga-Carlin M, Turner DJ, Siddiqui T, et al. Impact of the COVID-19 pandemic on diagnosis of new cancers: a national multicenter study of the veterans affairs healthcare system. Cancer. 2022;128(5):1048–56.
17. Ekberg S, Molin D, Pahnke S, Bergström F, Brånvall E, Smedby KE, et al. Impact of the COVID-19 pandemic on lymphoma incidence and short-term survival – a Swedish lymphoma register study. Acta Oncol. 2024;9(63):164–8.
18. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. Am J Epidemiol. 2017;186(9):1026–34.
19. Cerhan JR, Maurer MJ, Link BK, Feldman AL, Habermann TM, Jaye DL, et al. The lymphoma epidemiology of outcomes cohort study: design, baseline characteristics, and early outcomes. Am J Hematol. 2024;99(3):408–21.
20. Victorian Cancer Biobank. The Largest Consortium of Tissue Banks [Internet]. [cited 2022 Nov 22]. Available from: https://viccancerbiobank.org.au/
21. Wales Cancer Biobank. Biobank Cancer Cymru [Internet]. [cited 2023 May 19]. Available from: https://walescancerbank.com/

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Forsgren E, Ekberg S, Smedby KE, Nylund P, Sjöblom T, Flogegård M, et al. Evaluation of coverage, generalisability and validity of the U-CAN lymphoma biobank in Sweden: A comparison with nationwide registers. Br J Haematol. 2024;205(5):1794–1803. https://doi.org/10.1111/bjh.19732