ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO REAL WORLD DATA AND DIGITAL ONCOLOGY

SHORT COMMUNICATION

Understanding return to work after breast cancer through real-world data linkage framework: a Luxembourg digital oncology use case

B. Lima[1,2], A. Mafra[1,2], M. Schnell[1,3], L. Prévotat[4], M. Georges[5], N. Majery[5], S. Couffignal[1,3] & C. Backes[1,2,3]*

[1]Registre National du Cancer du Luxembourg (RNC), Strassen; [2]Cancer Epidemiology and Prevention Group (EPICAN), Department of Precision Health (DoPH), Luxembourg Institute of Health (LIH), Strassen; [3]Public Health Expertise (PHE), Department of Precision Health (DoPH), Luxembourg Institute of Health (LIH), Strassen; [4]Data Protection Office, Luxembourg Institute of Health (LIH), Strassen; [5]Service de Santé au Travail Multisectorel (STM), Luxembourg, Luxembourg

Available online xxx

Background: Breast cancer (BC) has a substantial impact on women of working age, making return to work (RTW) a critical outcome to evaluate. Research on RTW relies on individual-level real-world data integrating clinical and occupational information. Luxembourg offers a unique setting for such studies, with a young population and a high number of crossborder workers, creating specific challenges and opportunities for studying RTW. This study aims to describe a secure data linkage framework integrating population-based cancer registry, occupational health records, and questionnaire data to evaluate RTW after BC in Luxembourg.

Materials and methods: A cross-sectional pilot study was conducted linking data from three sources: from Luxembourg's National Cancer Registry (RNC), occupational health records from the Service de Santé au Travail Multisectoriel (STM), and results from a 27-item self-administered questionnaire. Eligible women were diagnosed with BC between 2013 and 2018, were younger than 65 years, employed under Luxembourgish labour law, and followed by STM. Potential participants were identified through STM records and invited to complete an electronic questionnaire and provide explicit consent for data linkage via REDCap® (Research Electronic Data Capture). A deterministic linkage between RNC and STM was enabled by using a common pseudonymisation service. Sociodemographic, clinical, and work-related characteristics were compared between linked and nonlinked participants and between women who did and did not RTW.

Results: Among 452 women contacted, 171 completed the questionnaire and provided informed consent (38% response rate). Ninety-six were successfully linked with the RNC dataset. Having Luxembourg as nationality or country of residence was associated with successful linkage ($P < 0.001$). Among linked participants, 74 (77.1%) RTW after BC-related leave. Women who RTW were significantly younger at diagnosis than those who did not, and most resumed work in the same sector and company.

Conclusions: Pseudonymised deterministic data linkage across three real-world data sources, including population-based cancer registry data, is feasible, and provides unique insights into RTW after BC. Crossborder residency patterns and heterogeneous cancer definitions between sources were key barriers to complete linkage. The population-based approach enables a more complete picture of RTW trajectories among women with BC, supporting robust real-world evidence and informing RTW surveillance and occupational health policies in highly mobile European crossborder regions.

Key words: data linkage, population-based cancer registry, return to work

INTRODUCTION

Breast cancer (BC) is the most diagnosed cancer among women in Europe, with more than half of new cases occurring before the age of 65 years.[1,2] As survival improves, growing numbers of women live with the long-term consequences of BC and its treatment.[3] Among women of working age, return to work (RTW) has become an important survivorship outcome, influencing quality of life, financial security, social participation, and workplace productivity.[4,5]

Comprehensive evaluation of RTW requires individual-level, real-world data integrating clinical, occupational and patient-reported information. Population-based cancer registries (PBCRs) provide high-quality cancer information,

*Correspondence to: Dr Claudine Backes, Cancer Epidemiology and Prevention Group (EPICAN), Department of Precision Health, Luxembourg Institute of Health, 1A-B, Rue Thomas Edison, L-1445 Strassen, Luxembourg. Tel: +352 621540467

E-mail: Claudine.Backes@lih.lu (C. Backes).

2949-8201/© 2026 The Authors. Published by Elsevier Ltd on behalf of European Society for Medical Oncology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

with one of their primary aims being cancer surveillance and the evaluation of quality of care. PBCR systematically collect information on all newly diagnosed or treated cancer cases within a well-defined geographical area. Assessing complementary dimensions such as occupational health and survivorship would require additional data sources, including occupational health records and patient-reported questionnaires. Linking these data sources enables a more comprehensive assessment of RTW than any single source alone.[6,7] Nevertheless, crossinstitutional data linkage remains challenging because of governance, data protection, and technical requirements.[8] Although frameworks have been proposed to guide such initiatives,[9] practical examples of their implementation within routine cancer care remain limited.

With a highly international labour market, in which nearly half of employees are crossborder workers,[10] Luxembourg offers a unique setting for studying RTW in population-based approaches. Luxembourg's National Cancer Registry (RNC) is a clinical PBCR that systematically collects and standardises data on all new cancer cases diagnosed and/or treated in Luxembourg. Within the RELIANCE project (REaL-life cANCEr epidemiology to identify risk factors for Breast Cancer with particular focus on prevention and care), established under the second National Cancer Plan, the RELIANCE-BC pilot study was developed to assess BC quality of care and survivorship in Luxembourg. A collaboration with the Service de Santé au Travail Multisectoriel (STM), Luxembourg's largest occupational health service provider, was established, which covers >60% of the national workforce. In Luxembourg, all employers must be affiliated with an occupational health service. Employees without a sector-specific service are automatically covered by STM, which provides occupational health follow-up across multiple economic sectors. The STM plays an important role in the RTW process, as prolonged sickness absence requires occupational health assessment before employees resume work. The STM provides complementary information on employees working under the Luxembourg labour law. This structured system provides a unique opportunity to integrate clinical and occupational data to study RTW after BC. The linkage of PBCR and occupational health data offers both opportunities and methodological challenges in a highly mobile workforce.

This study aimed to describe the governance, legal, and technical frameworks developed to securely integrate PBCR data, occupational health, and patient-reported data for studying RTW after BC in Luxembourg. We further illustrate the characteristics and analytical value of the linked data, providing practical insights for future real-world evidence generation using RNC data linkage.

MATERIALS AND METHODS

The RELIANCE BC pilot study is a cross-sectional observational study investigating RTW after BC through linking data from three complementary sources: RNC data, STM occupational health records, and a 27-item self-administered questionnaire.[11] The patient-reported questionnaire captured sociodemographic and socioeconomic factors, work history, social support, medical information, lifestyle factors, and elements of the Work Ability Index.[12] The RNC provided standardised and validated clinical information, including BC characteristics and treatments of all new cancer cases.

This study comprised women diagnosed with BC between January 1, 2013, and December 31, 2018, who were employed, were younger than 65 years (the legal retirement age in Luxembourg) at diagnosis, and had undergone an STM examination. Consequently, a total of 452 potential participants were identified through the STM's occupational health records.

The questionnaire was created in four languages and implemented in REDCap® (Vanderbilt University, Nashville, TN), a secure, web-based platform for research data collection.[13,14] To ensure data integrity, unique digital access keys (tokens) were generated to identify the blank questionnaires assigned to potential participants. The STM mapped these tokens to potential participants in their database via a correspondence table (first correspondance table) and mailed 452 invitation letters. These personalised tokens linked each invitee to a single blank questionnaire, preventing duplicate submissions, tracking response rates, and securing individual access. Each invitation contained information about the study, including a personalised quick response (QR) code that provided secure access to the informed consent form, data protection notices, and the respective questionnaire. The data protection notice described how personal data would be gathered, stored, used, and protected. Only participants who gave their explicit consent for participation and data linkage were included.

The linkage procedure comprised five sequential steps (Figure 1). Firstly, the STM identified eligible participants and distributed study invitations containing individualised REDCap® access tokens. Secondly, participants completed the electronic informed consent and questionnaire in REDCap®. Thirdly, for consenting participants, the STM requested the generation of pseudonyms via the pseudonymisation service provided by the Luxembourg Institute of Health (LIH) and used by the RNC. Using a correspondence table tracking pseudonyms and tokens per participant, the STM dataset was securely transferred to the RNC within the LIH.

Fourthly, the RNC used the shared pseudonyms to extract validated clinical information from its database, including diagnosis, tumour characteristics, and treatments. Finally, STM occupational health data, RNC clinical data, and REDCap® questionnaire responses were combined (Supplementary Figure S1, available at https://doi.org/10.1016/j.esmorw.2026.100782) through a deterministic data linkage. In the dataset shared with researchers a study-specific identifier replaced the pseudonym. A third correspondence table tracked the connection between pseudonyms and these study identifiers.

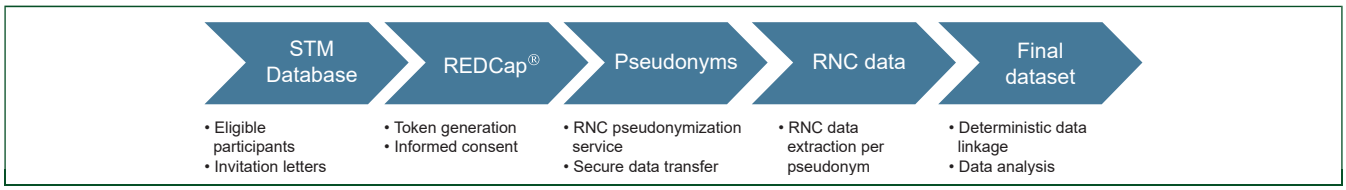

Figure 1. Data linkage process between three data sources. STM (Service de Santé au Travail Multisectoriel) provides occupational health records and manages participant recruitment; RNC (National Cancer Registry) provides clinical data; and REDCap® (Research Electronic Data Capture) captures self-administered questionnaire data.

Data processing followed EU General Data Protection Regulation (GDPR) 2016/679 requirements,[15] with oversight from the LIH and STM data protection officers (DPOs); a data protection impact assessment (DPIA) was carried out before starting data processing. Data protection was ensured through a privacy-by-design approach implemented throughout the project design. Potential privacy risks were systematically assessed and mitigated through appropriate technical and organisational measures, including pseudonymisation, access controls via role-based permissions, and data minimisation. This risk-based framework reduced the likelihood of participant reidentification. RTW status was obtained from the STM occupational health records, enabling two exploratory analyses. Firstly, to evaluate linkage feasibility and assess potential selection bias, sociodemographic, clinical, and occupational characteristics were compared between consenting participants successfully linked to the RNC and those who could not be linked. Secondly, among participants successfully linked to the RNC, sociodemographic characteristics were compared between women who returned to work and those who did not.

Continuous variables were summarised as means (± standard deviation) and compared using Student's *t*-test. Categorical variables were presented as counts and percentages and compared using the chi-square test or Fisher's exact test, as appropriate. Statistical significance was defined for a P value of <0.05. All analyses and data visualisations were carried out using R software (v.4.2.2, R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

A multistage linkage protocol was successfully implemented to integrate RNC, STM occupational health, and questionnaire response data while maintaining participant confidentiality (Figure 1). Among the 452 women identified through STM occupational health records, 171 (38%) provided informed consent (Figure 2). Of these, 96 participants were successfully linked to RNC BC data (linkage rate, 56%), comprising 90 with a first primary invasive BC, five with *in situ* carcinoma, and one with a prior diagnosis of another tumour type. Characteristics of consenting participants are presented in Table 1.

Comparing participants linked with RNC data ($n = 96$) with those not linked ($n = 75$), nationality and country of residence at diagnosis, were significantly associated with successful RNC linkage ($P < 0.001$ for both). Among linked participants, 71.9% were residents of Luxembourg, which is consistent with the RNC's main objective. In contrast, 24.7% of nonlinked participants had Germany as country of residence and 17.8% Luxembourg. Participants of German (20.5%), Belgian (11%), and French (43.8%) nationality were more frequent among nonlinked participants, whereas Luxembourgish and Portuguese nationality were

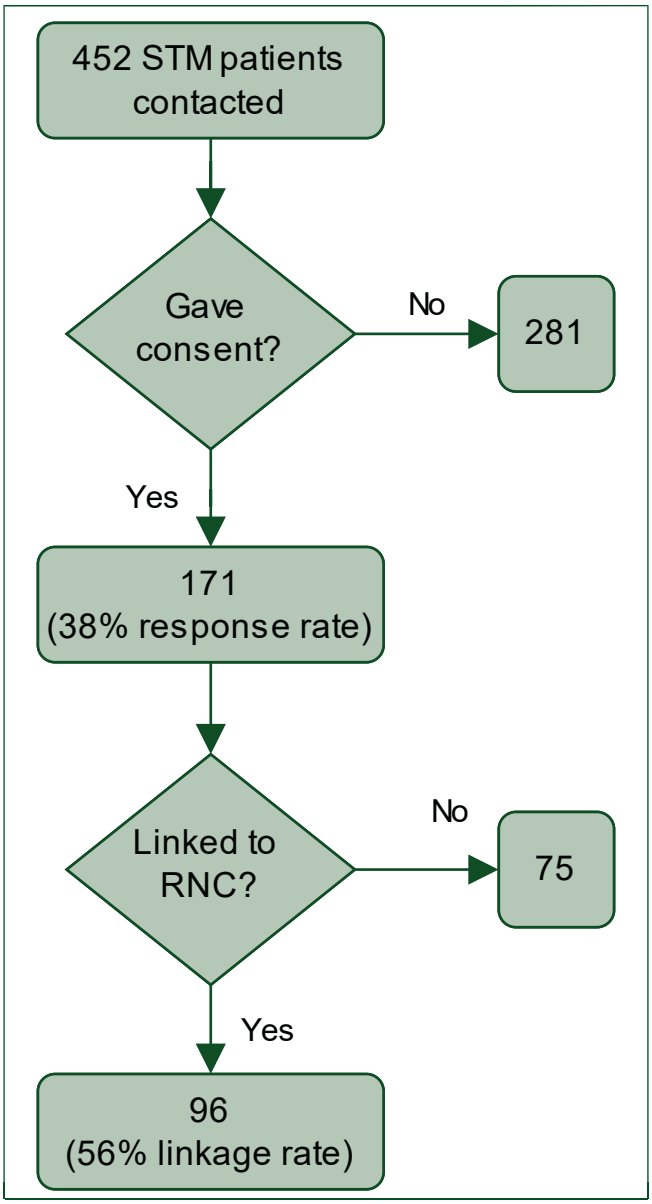

Figure 2. Flowchart with patients' recruitment and data linkage.

Table 1. Comparison of sociodemographic, clinical, and work-related variables between STM patients linked and not linked to RNC

Variable	Overall ($N = 171$)[a]	Not linked ($n = 75$)[a]	RNC ($n = 96$)[a]	P[b]
Age (years)	46.5 (±7.0)	46.3 (±7.6)	46.6 (±6.5)	0.77
Missing	2	2	0	
Nationality				<0.001
Luxembourgish	47 (27.8)	8 (11.0)	39 (40.6)	
German	16 (9.5)	15 (20.5)	1 (1.0)	
Belgian	15 (8.9)	8 (11.0)	7 (7.3)	
French	54 (32.0)	32 (43.8)	22 (22.9)	
Portuguese	22 (13.0)	5 (6.8)	17 (17.7)	
Other	15 (8.9)	5 (6.8)	10 (10.4)	
Missing	2	2	0	
Residence				<0.001
Luxembourg	82 (48.5)	13 (17.8)	69 (71.9)	
Germany	22 (13.0)	18 (24.7)	4 (4.2)	
Belgium	20 (11.8)	13 (17.8)	7 (7.3)	
France	45 (26.6)	29 (39.7)	16 (16.7)	
Missing	2	2	0	
BMI (kg/m^2)	25.2 (±5.1)	24.9 (±4.4)	25.4 (±5.5)	0.49
Missing	2	2	0	
Education				0.59
Bachelor degree and above	68 (39.8)	28 (37.3)	40 (41.7)	
Secondary and postsecondary	69 (40.4)	32 (42.7)	37 (38.5)	
Lower secondary	21 (12.3)	11 (14.7)	10 (10.4)	
Primary or less	13 (7.6)	4 (5.3)	9 (9.4)	
Income to your support				0.67
Very difficult	4 (2.3)	2 (2.7)	2 (2.1)	
Fairly difficult	45 (26.3)	20 (26.7)	25 (26.0)	
Quite easily	100 (58.5)	46 (61.3)	54 (56.3)	
Very easily	22 (12.9)	7 (9.3)	15 (15.6)	
Smoke	73 (42.9)	35 (46.7)	38 (40.0)	0.38
Missing	1	0	1	
Alcohol consumption				0.29
Never	30 (17.5)	18 (24.0)	12 (12.5)	
< once a month	31 (18.1)	15 (20.0)	16 (16.7)	
Once a month	29 (17.0)	13 (17.3)	16 (16.7)	
1-2 days a week, or only on weekends	65 (38.0)	23 (30.7)	42 (43.8)	
3-4 days a week	14 (8.2)	5 (6.7)	9 (9.4)	
Every day or almost	2 (1.2)	1 (1.3)	1 (1.0)	
Past occupation sector				0.18
Mental	71 (44.7)	23 (35.9)	48 (50.5)	
Mixed	65 (40.9)	31 (48.4)	34 (35.8)	
Physical	23 (14.5)	10 (15.6)	13 (13.7)	
Missing	12	11	1	
Present occupation sector				0.19
Mental	76 (53.5)	27 (45.0)	49 (59.8)	
Mixed	48 (33.8)	25 (41.7)	23 (28.0)	
Physical	18 (12.7)	8 (13.3)	10 (12.2)	
Missing	29	15	14	

BMI, body mass index; RNC, Registre National du Cancer; STM, Service de Santé au Travail Multisectoriel.
[a]Values are mean (±SD) or *n* (%).
[b]Two-sample *t*-test; Pearson chi-square test; or Fisher's exact test.

more common among linked participants (Table 1). Except for nationality and country of residence, no statistically significant differences were observed between linked and nonlinked participants.

Among the 96 linked participants, 74 (77.1%) RTW after BC-related medical leave (Table 2). Participants who RTW were significantly younger at diagnosis than those who did not RTW (43.6 ± 5.7 versus 48.7 ± 7.1, $P = 0.004$). Regarding career continuity, 87.7% resumed employment in the same sector, and 71.2% returned to the same company. The proportion of Luxembourgish nationals was lower among participants who RTW than that of those who did not (35.1% versus 59.1%).

Table 2. Comparison of sociodemographic variables between participants who returned to work and those who did not

Variable	Overall ($N = 96$)[a]	RTW no ($n = 22$)[a]	RTW yes ($n = 74$)[a]	P[b]
Age at diagnosis	44.8 (±6.4)	48.7 (±7.1)	43.6 (±5.7)	0.004
Past occupation sector				0.53
Mental	48 (50.5)	9 (40.9)	39 (53.4)	
Mixed	34 (35.8)	10 (45.5)	24 (32.9)	
Physical	13 (13.7)	3 (13.6)	10 (13.7)	
Missing	1	0	1	
Present occupation sector				
Mental			42 (56.8)	
Mixed			22 (29.7)	
Physical			10 (13.5)	
Missing			0	
Working in the same company			52 (71.2)	
Missing			1	
Working in the same sector			64 (87.7)	
Missing			1	
Working in the same job description			38 (52.1)	
Missing			1	
Nationality				0.57
Luxembourgish	39 (40.6)	13 (59.1)	26 (35.1)	
German	1 (1.0)	0 (0.0)	1 (1.4)	
Belgian	7 (7.3)	1 (4.5)	6 (8.1)	
French	22 (22.9)	3 (13.6)	19 (25.7)	
Portuguese	17 (17.7)	3 (13.6)	14 (18.9)	
Other	10 (10.4)	2 (9.1)	8 (10.8)	
Residence				0.24
Luxembourg	69 (71.9)	17 (77.3)	52 (70.3)	
Germany	4 (4.2)	2 (9.1)	2 (2.7)	
Belgium	7 (7.3)	0 (0.0)	7 (9.5)	
France	16 (16.7)	3 (13.6)	13 (17.6)	

RTW, return to work.
[a]Values are mean (±SD) or *n* (%).
[b]Welch Two-sample *t*-test; Fisher's exact test; or Pearson chi-square test.

DISCUSSION

This study demonstrated the feasibility of developing a novel Real World Data (RWD) linkage framework in Luxembourg. The linked dataset enabled an integrated assessment of clinical, occupational, and patient-reported. and patient-reported factors related to RTW after BC in a population-based cohort with the potential to reduce selection bias. Differences in the populations covered by the RNC and STM, particularly regarding cross-border workers, were the main factors associated with unsuccessful linkage. Furthermore, heterogeneous cancer definitions across data sources represented an additional challenge to the linkage

process, as they can lead to discrepancies in the populations captured by each source.

Among the successfully linked participants, most women returned to work after BC-related leave. Younger age at diagnosis was associated with RTW. Although these findings are exploratory, they illustrate the type of research questions that can be addressed by integrating registry, occupational health, and patient-reported data.

By combining all three data sources, the linkage framework generated a multidimensional dataset that is not available from any single source. Importantly, this framework extends the analytical value of PBCRs without requiring expansion of their routine data collection. Rather than collecting occupational or patient-reported information within the registry itself, secure linkage enables the integration of complementary data for research purposes. Establishing a robust governance framework was a key prerequisite for successful linkage. Legal permissions and data protection requirements were assessed by the DPOs of the participating institutions and a DPIA was completed to ensure GDPR compliance.[15] A major facilitator of the linkage was the use of a common pseudonymisation service by both RNC and STM. This enabled deterministic data linkage without exchanging direct identifiers while maintaining confidentiality and data security. Clearly defined data flows and institutional responsibilities facilitated collaboration between participating organisations.

A participation rate of 38% is comparable with those reported in previous RTW studies among cancer survivors.[16] Although participation bias cannot be excluded, this demonstrates the feasibility of engaging cancer survivors in studies combining population-based clinical, occupational, and patient-reported information. Incomplete linkage was mainly explained by differences in population coverage rather than technical limitations. The lower linkage rate among crossborder workers was expected given the RNC's and STM's inclusion criteria. Consequently, linkage success reflects not only technical performance but also differences in the populations captured by each data source. In addition, STM cancer information is self-reported, whereas RNC inclusion follows international PBCR standards and definitions. Differences in case ascertainment and diagnostic and treatment pathways may therefore contribute to incomplete linkage.

These findings highlight an important consideration for countries with mobile workforces: unsuccessful linkage does not necessarily indicate poor data quality but may reflect differences in the populations captured by each data source. Similar initiatives should therefore consider the scope, inclusion criteria, and underlying populations of contributing data sources when interpreting linkage success.

A major strength of this study is the broad population coverage. The STM is one of Luxembourg's largest occupational health providers and covers a substantial proportion of the national workforce, providing a valuable opportunity to study occupational trajectories after cancer. However, they do not have access to clinical health records. The RNC covers all newly diagnosed BC cases diagnosed within a specific period, supporting the reduction of selection biases.

Nevertheless, transferability of this framework will depend on local characteristics, including cancer registry coverage, occupational health structures, legal requirements and available pseudonymisation mechanisms. Therefore, successful implementation requires adaptation to local data ecosystems rather than direct replication of a single model. Beyond the Luxembourg setting, this framework illustrates how secure linkage of complementary real-world data sources can enhance the value of population-based cancer registries for survivorship research while preserving their core surveillance function. These findings support robust real-world evidence and informing RTW surveillance and occupational health policies in highly mobile European cross-border regions.

FUNDING

None declared.

DISCLOSURE

The authors have declared no conflicts of interest.

REFERENCES

1. European Union. *ECIS—European Cancer Information System*. Joint Research Centre. 2025. Available at: https://ecis.jrc.ec.europa.eu. Accessed May 2, 2026.
2. Europa Donna. *The European Breast Cancer Coalition*. Milan: Europa Donna. 2024. Available at: https://www.europadonna.org/breast-cancer/. Accessed May 23, 2026.
3. Gregorowitsch ML, van den Bongard H, Couwenberg AM, et al. Self-reported work ability in breast cancer survivors; a prospective cohort study in the Netherlands. *Breast*. 2019;48:45-53.
4. Richter-Ehrenstein C, Martinez-Pader J. Impact of breast cancer diagnosis and treatment on work-related life and financial factors. *Breast Care (Basel)*. 2021;16(1):72-76.
5. Islam T, Dahlui M, Majid H, Nahar A, Taib N, Su T. Factors associated with return to work of breast cancer survivors: a systematic review. *BMC Public Health*. 2014;14(Suppl 3):S8.
6. Tham N, Skandarajah A, Hayes IP. Colorectal cancer databases and registries in Australia: what data is available? *ANZ J Surg*. 2022;92(1-2):27-33.
7. Pearson C, Fraser J, Peake M, et al. Establishing population-based surveillance of diagnostic timeliness using linked cancer registry and administrative data for patients with colorectal and lung cancer. *Cancer Epidemiol*. 2019;61:111-118.
8. Heins MJ, de Ligt KM, Verloop J, Siesling S, Korevaar JC. Opportunities and obstacles in linking large health care registries: the primary secondary cancer care registry—breast cancer. *BMC Med Res Methodol*. 2022;22(1):124.
9. Lima B, Hasan F, Kannan P, et al. Optimizing data linkage for maximizing the potential of Luxembourg's national cancer registry: a comprehensive scoping review. *Front Oncol*. 2025;15:1679408.
10. STATEC—Institut national de la statistique et des études économiques. *Panorama of the Luxembourg Labour Market on May 1st*. Luxembourg: STATEC; 2025.
11. Mafra A, Talbi M, Cassells G, et al. Returning to work after breast cancer diagnosis: a cross-sectional study of socioeconomic and clinical factors associated with return to work in Luxembourg. *BMC Womens Health*. 2026;26(1):107.

12. Ilmarinen J. The Work Ability Index (WAI). *Occup Med*. 2006;57(2):160.
13. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208.
14. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381.
15. General Data Protection Regulation. *Regulation (EU) 2016/679* 2016.
16. Grunfeld EA, Rixon L, Eaton E, Cooper AF. The organisational perspective on the return to work of employees following treatment for cancer. *J Occup Rehabil*. 2008;18(4):381-388.