The completion and concordance of cancer data collected on-site and from central registries in England and Wales

Ashley Jackson[1], Pradeep S Virdee[2], Sharon Tonner[2], Jason Oke[2], Rafael Perera[2], Kaveh Riahi[3], Ying Luan[4], Sara Hiom[3], Harpal Kumar[3], Harit Nandani[4], Kathryn N Kurtzman[4], FD Richard Hobbs[2], Mark R Middleton[1], Brian D Nicholson[2]

[1]Department of Oncology, University of Oxford, Oxford, UK
[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
[3]GRAIL Bio UK, GRAIL, London, UK
[4]GRAIL, Menlo Park, CA, USA

Background & Aims

- Cancer is one of the leading causes of death worldwide, leading to significant resources being committed to cancer research[1,2]
- Using central registries is one possible method for improving the resource burden associated with cancer research [3,4]
- Investigating the completeness, concordance, and timeliness of data collected both on-site and from central registries is necessary to understand the feasibility of using these resources [5,6] (Fig 1.1)

Figure 1.1. Characteristics of data quality of cancer databases.

- We aimed to compare cancer data obtained on-site during a prospective cohort study in England and Wales with data from English and Welsh national cancer registries to examine data completeness, concordance, and timeliness

Methods

- On-site data were collected during a prospective cohort study in England and Wales (SYMPLIFY)
- Linked central data were obtained from Digital Health and Care Wales (DHCW), the Welsh Cancer Intelligence and Surveillance Unit (WCISU), the English National Cancer Registration Dataset (NCRD), and the English Rapid Cancer Registration Dataset (RCRD)
- Data cuts from these datasets were retrieved regularly between Apr 2022 and Sep 2023 (Fig 2.1)
- Four data fields were investigated: ICD-10 code (cancer site), ICD-O-3 code (morphology), overall clinical stage, and TNM classification

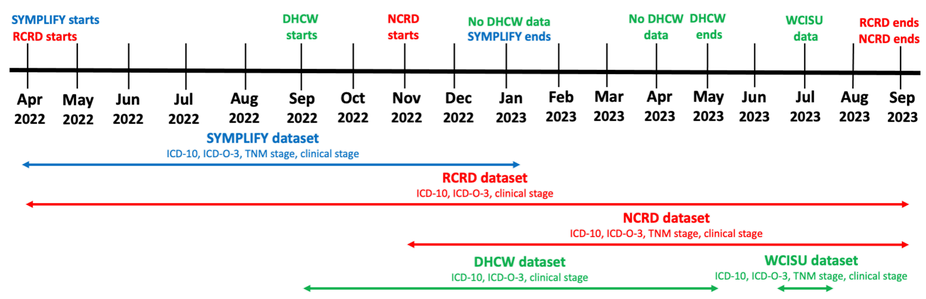

Figure 2.1. Timeline of the data cuts and data fields available for each dataset.

Completeness

There were 259, 121, 226, 291, 122, and 112 cancers recorded in SYMPLIFY-England, SYMPLIFY-Wales, RCRD, NCRD, DHCW, and WCISU, respectively, by Sep 2023. ICD-10 completion was 100% at the final time point for each dataset, while completion for ICD-O-3, clinical stage, and TNM stage varied among datasets (Table 3.1).

Table 3.1. Proportion (%) of cancers reported at the final timepoint for each dataset with complete data for ICD-10, ICD-O-3, TNM stage, and clinical stage.

Dataset	ICD-10	ICD-O-3	TNM	Clinical Stage
SYMPLIFY - England	100	98.5	82.6	100
SYMPLIFY - Wales	100	100	73.6	100
RCRD	100	100	-	72.6
NCRD	100	100	76.3	99.7
DHCW	100	83.6	-	43.4
WCISU	100	100	76.8	100

By the final data cut for RCRD and NCRD, 76.8% and 91.1% of the SYMPLIFY-England cancers were reported in the registries, respectively. Similarly, 87.6% and 81.0% of the SYMPLIFY-Wales cancers were reported in DHCW and WCISU, respectively, by the final time point for each dataset.

Table 3.2. Date of diagnosis for cancers found in (a) NCRD (n=55) and (b) WCISU (n=14) at the final registry data cut available (Sep 2023 and Jul 2023, respectively) that were not reported in the final SYMPLIFY data cut (Jan 2023).

a) NCRD

Time Frame	Number of cancers diagnosed n (%)	Cumulative diagnoses n (%)
< 3 months post-enrolment	30 (54.5)	30 (54.5)
3-6 months post-enrolment	17 (30.9)	46 (85.5)
6-9 months post-enrolment	8 (14.5)	55 (100)

b) WCISU

Time Frame	Number of cancers diagnosed n (%)	Cumulative diagnoses n (%)
< 3 months post-enrolment	6 (42.9)	6 (42.9)
3-6 months post-enrolment	5 (35.7)	11 (78.6)
6-9 months post-enrolment	3 (21.4)	14 (100)

There were 55 and 14 cancers registered in NCRD and WCISU, respectively, at the final time point that were not recorded in the SYMPLIFY dataset. Of these, 54.5% and 42.9% were diagnosed within the 3-month post-enrolment mandatory follow-up period in SYMPLIFY (Table 3.2).

Concordance

Concordance between SYMPLIFY and the central registries was highest for ICD-10 and broad ICD-O-3 morphology groupings but was lower for ICD-O-3 code, clinical stage, and TNM staging (Fig 3.1).

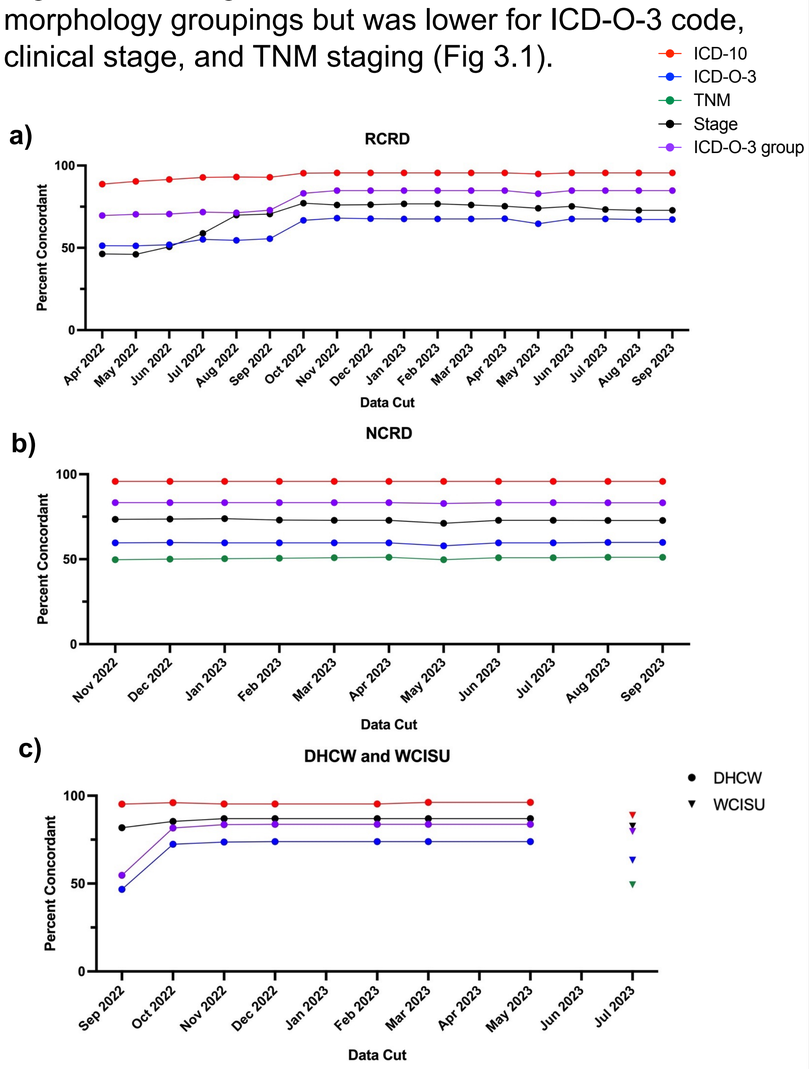

Fig 3.1. Concordance between data fields in SYMPLIFY and (a) RCRD, (b) NCRD, and (c) DHCW and WCISU.

Overall TNM stage concordance was low between SYMPLIFY and NCRD (51.1%), but there was higher concordance for individual T, N, and M stages (Table 3.3).

Table 3.3. Concordance (%) of T stage, N stage, M stage, and overall TNM stage for each cancer recorded in both SYMPLIFY and the NCRD dataset at the final timepoint in Sep 2023.

Variable	Concordance
T stage	73.9
N stage	78.4
M stage	90.9
TNM stage	51.1

Timeliness

SYMPLIFY reached completeness of cancer data in Nov 2022, approximately 12 months after the study recruitment ended. NCRD and RCRD both reached completeness of cancer registrations compared to the corresponding final data cuts at 13 months post-enrolment, while DHCW reached completeness of cancer registrations at 15 months post-enrolment (Fig 3.2).

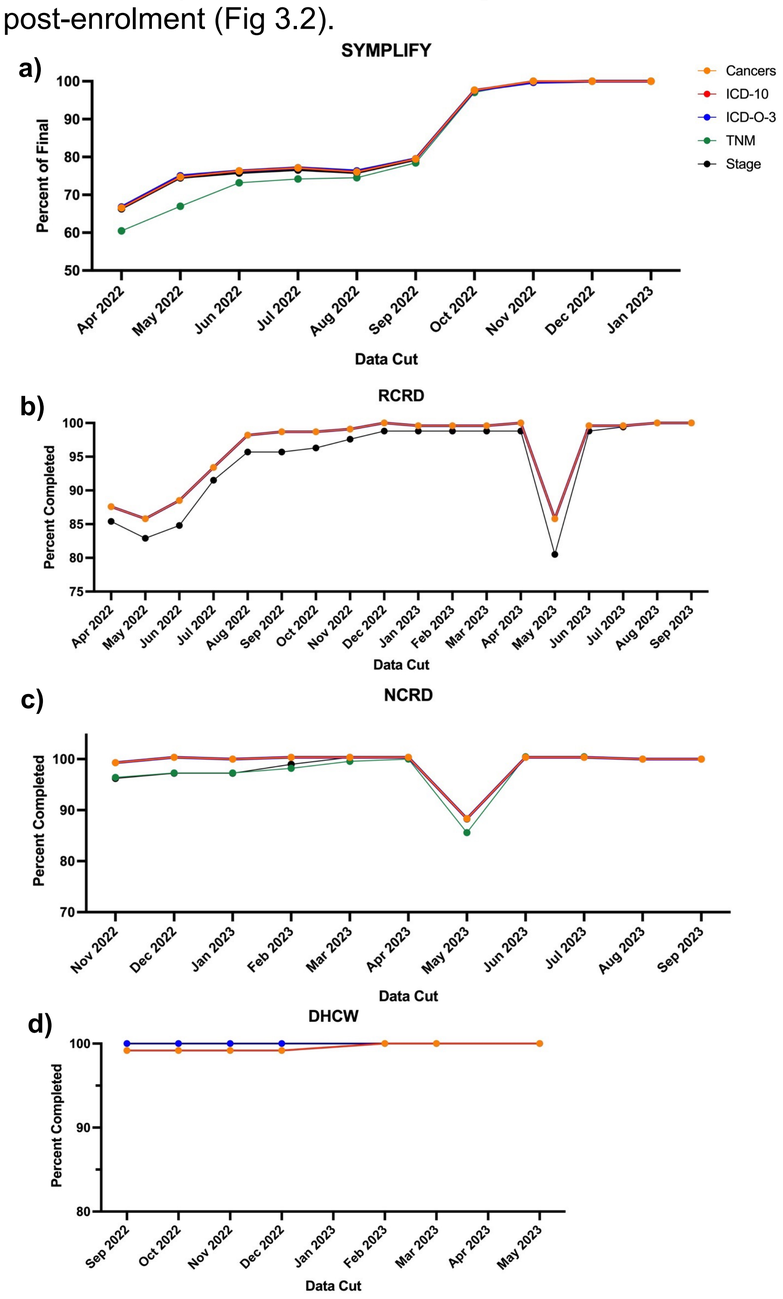

Figure 3.2 Timeliness of completeness of data fields for (a) SYMPLIFY, (b) RCRD, (c) NCRD, and (d) DHCW data compared to the final dataset for each data source. Due to a data issue at one of the hospital sites, there was a temporary drop in cancers in the RCRD and NCRD datasets in May 2023.

Discussion & Conclusions

Summary

- To our knowledge, this is the first study to compare the completeness, concordance, and timeliness of cancer diagnosis data collected at study sites in England and Wales with centrally collected registry data
- We demonstrate comparable completeness, accuracy, and timeliness between on-site and central registry cancer data
- We demonstrate the potential advantages of using registry data in cancer research, which may allow for a more complete and thorough follow-up of patients

Next Steps

- Future studies should evaluate the role of patient and cancer characteristics on the completion, concordance, and timeliness of cancer data in these datasets (Fig 4.1)

Conclusion

- Our study contributes to ongoing research which supports the use of centralised registries in decentralised cancer research

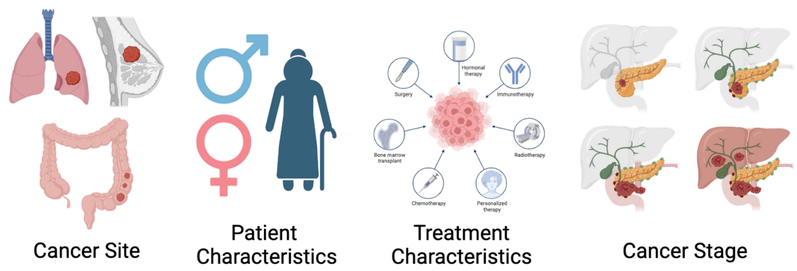

Figure 4.1. Characteristics known to impact the completion and accuracy of cancer registry data [7-10].

References

1. How we spend your money Cancer Research UK [cited 2023 May 17].
2. Emanuel EJ, Schnipper LE, Kamin DY, Levinson J, Lichter AS. The costs of conducting clinical research. J Clin Oncol. 2003;21(22):4145-50.
3. Anderson BR, Gotlieb EG, Hill K, McHugh KE, Scheurer MA, Mery CM, et al. Registry-based trials: a potential model for cost savings? Cardiol Young. 2020;30(6):807-17.
4. Li G, Sajobi TT, Menon BK, Korngut L, Lowerison M, James M, et al. Registry-based randomized controlled trials- what are the advantages, challenges, and areas for future research? J Clin Epidemiol. 2016;80:16-24.
5. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer. 2009;45(5):747-55.
6. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. Eur J Cancer. 2009;45(5):756-64.
7. Seneviratne S, Campbell I, Scott N, Shirley R, Peni T, Lawrenson R. Accuracy and completeness of the New Zealand Cancer Registry for staging of invasive breast cancer. Cancer Epidemiol. 2014;38(5):638-44.
8. Søgaard M, Olsen M. Quality of cancer registry data: completeness of TNM staging and potential implications. Clin Epidemiol. 2012;4 Suppl 2(Suppl 2):1-3.
9. Korhonen P, Malila N, Pukkala E, Teppo L, Albanes D, Virtamo J. The Finnish Cancer Registry as follow-up source of a large trial cohort--accuracy and delay. Acta Oncol. 2002;41(4):381-8.
10. Larsen IK, Småstuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. Eur J Cancer. 2009;45(7):1218-31.

Acknowledgements

SYMPLIFY was funded by GRAIL Bio UK, with support from National Health Service (NHS) England, NHS Wales, the National Institute for Health Research (NIHR), and Oxford NIHR Biomedical Research Centre. MRM and BN are supported by the NIHR Biomedical Research Centre at Oxford and Cancer Research UK. AJ is supported by the Rhodes Trust. The views expressed in this poster are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. We thank Peter Johnson, Lennard Lee, and Tom Crosby at NHS England and NHS Wales for their assistance and insights over the course of the study. We also thank all patients who participated, staff, and investigators.