Whitepaper: NSW Clinical Cancer Registry Data 2008 - 2012

An overview and guide for prospective researchers

Version 1.4 | 22 October 2021

NSW GOVERNMENT

Cancer Institute NSW

Contents

Abstract

The purpose of this paper is to briefly articulate some background on the coverage offered by the 2008 to 2012 Clinical Cancer Registry (ClinCR) data and the implications for designing a study based on these data.

It does not aim to inform on available data items or the quality and completeness of these data items (these are available elsewhere).

Definitions

Epidemiology - The study of the distribution of disease and determinants of health-related states or events in specified human populations and the application of this study to the control of human health problems.

Clinical epidemiology - The science of making predictions about individual patients by counting clinical events in similar patients, using strong scientific methods for studies of groups of patients to ensure that the predictions are accurate.

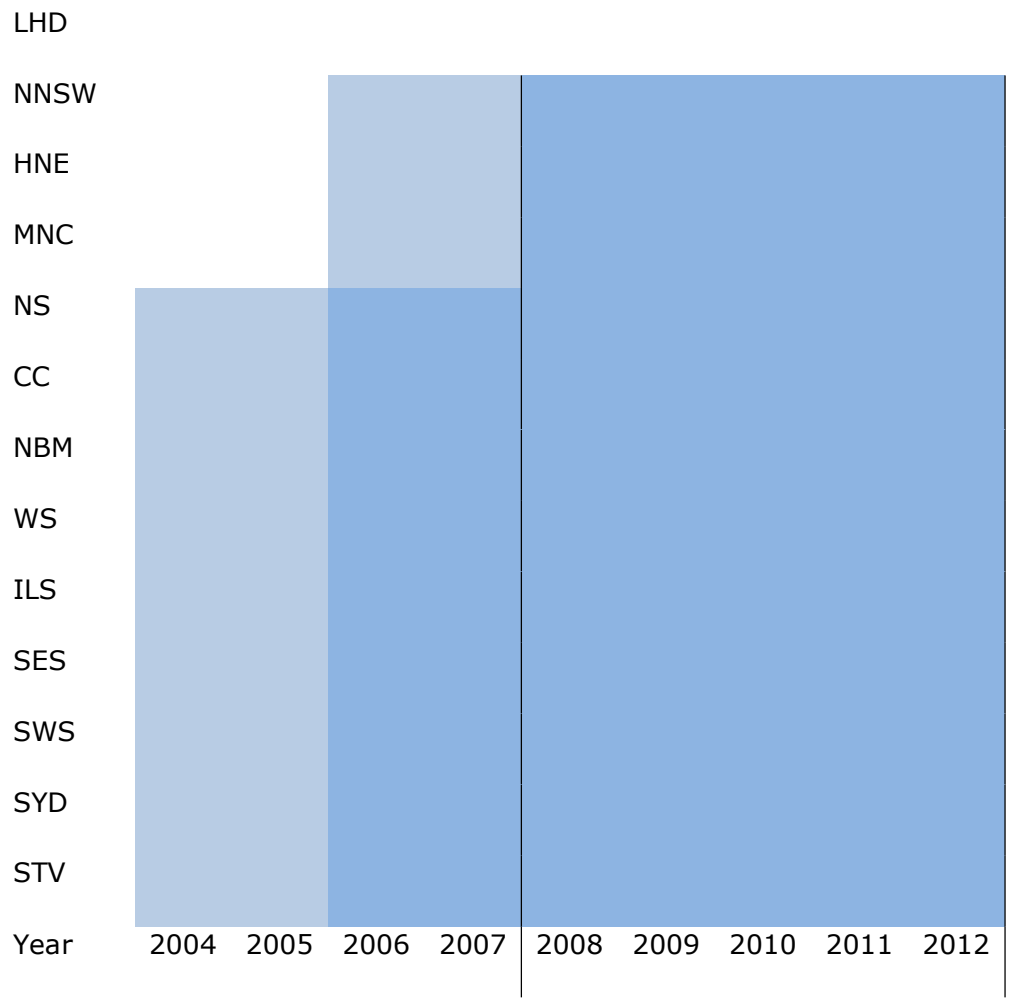
Health service utilisation - The measure of the population's use of the health care services available to them.

1.0 Unlinked Clinical Cancer Registries (ClinCRs)

The ClinCRs record basic patient and cancer data along with clinical data relating to episodes of care in public facilities within 8 Local Health Districts (LHDs) in NSW.

1.1 Coverage – Local Health Districts (LHDs) & Time

Initially 9 LHDs contributed data with 3 more joining in 2006. Data from any LHD in the first 2 years are considered test quality only, thus Jan 2008 is considered a useful and valid starting point by many researchers. Data are collected on new episodes of care to Dec 2012 and includes treatments for these episodes to 2014



1.2 Coverage –Episode types

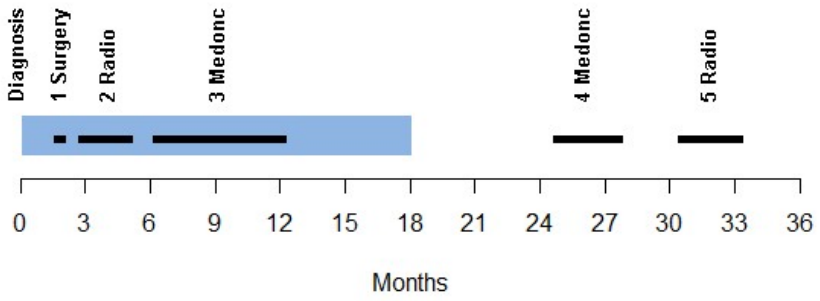
Episodes of care recorded in ClinCRs include:

Surgery
Medical Oncology & Haematology
Radiotherapy
Diagnostic
Admitted/Other

1.3 Coverage –Episode timing

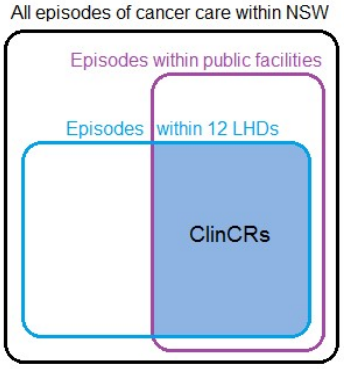
Generally only initial episodes of care occurring within a period of 18 months after diagnosis are recorded in ClinCRs. Studies focussing on long term treatment patterns, progression and recurrence will be difficult using ClinCRs data.

Example: episodes 4 & 5 below are unlikely to be recorded in ClinCRs.



1.4 Coverage – Location of episode of care

ClinCRs is not a population based registry[1]. It represents a specific subset of all episodes of care in NSW depending on the type (public private) and location (LHD) of the facility where care was received.



1.5 Death

The data field indicating vital status "xxx" is not to be used and has been suppressed.

1.6 Making sound inferences from ClinCRs

It should be clear ClinCRs does not have population coverage. Under no circumstances should it be considered a census or assumed to be a representative sample.

The availability and use of private facilities varies among LHDs. Patients can flow in and out of neighbouring LHDs utilising a mix of private and public facilities during their overall course of treatment. All these factors will affect the capture and completeness of coverage of a person's total clinical data in the ClinCRs as they undergo treatment.

As a result of this design constraint considerable thought and great care is necessary when designing a valid study in order to make sound inferences and useful conclusions.

Example: Consider the clinical data of 3 people diagnosed with breast cancer in the ClinCRs.

Person ID	Episodes of care		
P1	Surgery	MedOnc	Radio
P2	Surgery	Radio	
P4	Surgery		

Person P1 has 3 episodes of care recorded. This may not represent their treatment in its entirety. It is possible additional episodes of chemotherapy were undertaken in a private facility and thus not recorded in the ClinCRs.

Person P2 has 2 episodes of care recorded. It is not clear if they either i) did not require adjuvant chemotherapy or ii) they required but did not receive any adjuvant chemotherapy or finally iii) if they did receive adjuvant chemotherapy but in a private facility.

Person P4 has only 1 episode of care recorded. This does not necessarily indicate absence of adjuvant care. Strong conclusions about evidence of incomplete or inadequate care from the ClinCRs are entirely unwarranted. It remains unclear from these data whether they received adjuvant treatment in a non-included LHD, or in a private facility, or did not receive any adjuvant treatment at all.

1.7 Health service utilisation studies

Health service utilisation studies are very challenging using ClinCRs data owing to the incomplete coverage described in section 1.4. Successful applications would need to demonstrate an understanding of the limitations in the ClinCRs target population & sampling frame and demonstrate a means to mitigate the inherent selection biases in order to make useful inferences. Failing to do this will result in studies with very poor external validity.

Example: a study trying to determine the proportion of women who received chemotherapy for a diagnosis of breast cancer might find a low proportion of women receiving adjuvant chemotherapy in general and also considerable variation among neighbouring LHDs. Rather than reflexively attributing these results to under-treatment and un-warranted variation there are in fact many competing, plausible hypotheses. The apparent low proportion women receiving of adjuvant chemo is possibly due to women receiving chemotherapy in private facilities. The variation may reflect the variation in private facilities available to women in different LHDs.

Successful study applications clearly demonstrate an understanding of the inherent sampling frame of the ClinCRs, its associated strengths & limitations and tailor their study aims & objectives accordingly.

Applications seeking to evaluate (and benchmark) "patterns of care" or "appropriateness of care" or "variation in patterns of care" without acknowledging the inherent difficulties and potential biases involved will result in studies with very poor external validity and do not have a good track record of approval.

1.8 Patterns of treatment within public facilities

The ClinCRs can offer insights into how treatments are used within and among public facilities.

Example: Delaney et al (2016) report on the use of hypofractionation for node negative breast cancer in public radiotherapy facilities in NSW 2008-12. They also examine relationships between the likelihood of hypofractionation treatment and age, laterality and other factors as well as examining the uptake of hypofractionation within facilities over time.

These studies seek to gain insights into the patterns of care delivered by specific (public) facilities. ClinCRs offers good coverage for episodes of care provided by public facilities over the period 2008-12.

1.9 Aetiological studies

The ClinCRs can offer insights into hypothesised causal associations for treatments.

Example: a study investigating if the risk of cisplatin associated admission for neutropenia is more common in the initial 4 weeks of treatment or the final 4 weeks. Additional relationships such as the likelihood of admission for neutropenia for different age groups undergoing treatment with cisplatin can be further studied.

These aims seek to gain insights into specific mechanisms relating to a treatment and are far less reliant on a population level, or representatively sampled data.

2.0 Linked Clinical Cancer Registries (L-ClinCRs)

The ClinCRs are able to be linked to the NSW Central Cancer Registry (CCR) and other datasets by the NSW Centre for Health Record Linkage (CHeReL).
ref http://www.cherel.org.au/master-linkage-key

Linking to the CCR can provide useful additional information regarding the person and their cancer. Importantly it can also provide additional insights into the coverage provided by ClinCRs for the study cancer(s) in question (e.g. patient P3 below was diagnosed with breast cancer however none of their subsequent episodes of care were recorded in the ClinCRs).

Patient ID	ClinCRs - Episodes of care			CCR - Cancer information		
P1	Surgery	MedOnc	Radio	C50	additional	info
P2	Surgery	Radio		C50	additional	info
P3				C50	additional	info
P4	Surgery			C50	additional	info

Linking to the population level NSW Admitted Patients Data Collection (APDC) is a powerful means to enrich the ClinCRs' surgical data.

2.1 Linkage and approvals

Linking to additional sources of data is a powerful means to enrich the ClinCRs datasets and broaden the scope of a study however there is an additional, necessary overhead of applications and approvals.

3.0 Notes for prospective researchers

Proposed studies naively treating the ClinCRs as a population based registry, or assuming it to be a representative sample, have a poor track record of approval.
"We aim to estimate the proportion of patients who get chemotherapy for breast cancer in NSW"

Proposed studies acknowledging potential biases and attempting to mitigate these with complex statistical methods also have a poor track record of approval.
"We aim to estimate the proportion of patients who get chemotherapy for breast cancer in NSW, potential bias as a result of people undergoing treatment in the private sector will be mitigated by a synthesis of weighting and post stratification based on a Bayesian hierarchical model…"

Proposed studies acknowledging potential biases and narrowing the scope of their study aims, objectives and inferences in recognition of these have a good track record of approval.
"We aim to study the changing patterns in chemotherapies for breast cancer in public hospitals over the period 2008-12. We are primarily interested in within-hospital changes over time of the index chemotherapy…"

"We aim to study the duration of the index chemotherapy treatment for patients undergoing chemotherapy treatment in a public hospital over the period 2008-12."

Statistical and scientific generalizations

When designing a study based on ClinCRs data it can be useful to consider the distinction between statistical and scientific generalisations.

A statistical generalisation involves the extrapolation of results from a sample to its source (or a future) population/scenario. A scientific generalisation is the process of constructing a correct statement about the way nature works[2] (i.e. deducing cause and effect).

Example: A valid statistical generalisation about the prevalence of drink driving among students would require observing the habits of a representative sample of students. However measuring the deleterious impact of drinking on driving ability (a scientific generalisation) does not necessarily require a representative sample; this can in fact be reliably ascertained from a convenience sample of students.

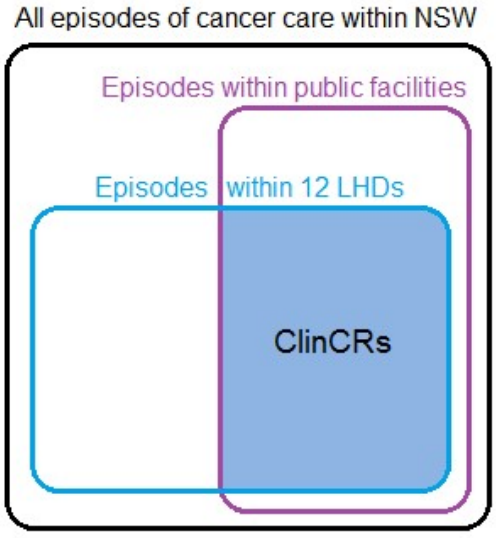
A study which seeks to determine the proportion of all people in NSW with bladder cancer who receive cisplatin chemotherapy (a statistical generalisation) would be challenging with ClinCRs data alone. However a study looking at the proportion of people in NSW with bladder cancer treated with cisplatin chemotherapy that have a related admission for febrile neutropenia (cause and effect) with ClinCRs data linked to hospital admission data is less sensitive to needing a representative sample and is a more feasible study aim.

4.0 Summary

Most successful study applications understand the coverage provided by the ClinCRs and tailor their study aims accordingly in order to make valid, and useful, conclusions.

Most successful study applications understand that missing data may reflect absence of a particular episode of care or alternatively, that care being delivered in the private sector or in a rural LHD.

Most successful study applications recognize the potential for major sources of bias in all observational data and give serious thought and consideration to these in their study design.



5.0 References

1 Olsen J et al; What is a population-based registry?
Scand J Public Health January 1999 vol. 27 no. 1 78

2 Rothman, K; Six Persistent Research Misconceptions
J Gen Intern Med. 2014 Jul;29(7):1060-4