The application and feasibility of using routine data sources for long term cancer clinical trial follow-up

Introduction

As cancer-specific survival rates increase, patients are living longer and therefore their chances of developing late recurrences and long term side-effects have also increased. This recurrence and toxicity data is important for developing the optimal management for patients, and also for providing information on long-term consequences of treatments as part of the informed consent process.

Historically, most radiotherapy trials have endeavoured to follow up patients for at least 10 years. The need to capture long-term outcome events is clear, but is especially pertinent with radiotherapy trials as adverse effects can continue to occur and develop years and sometimes decades after completion of treatment.

Obtaining prolonged clinical trial follow-up is becoming increasingly challenging, especially against the background of rising survival rates. In addition, workforce changes in cancer care have taken place following national guidance such as the 2016 NHS England publication “Achieving World Class Cancer Outcomes: Taking the strategy forward”(1). This document states that the uptake of risk stratified follow up should increase by 2017 and gives the following example: “We know that risk stratified follow up pathways for breast cancer patients can not only improve care for patients after treatment, but create more efficient pathways in the NHS that can, for example, reduce unnecessary outpatient appointments”. The fundamental need to obtain long term outcome data within a clinical trial, is therefore slightly at odds with the move towards risk stratified follow up. Hospitals who have already streamlined follow up may consider additional clinic visits for clinical trial assessment as a financial burden and some patients view extra hospital visits negatively.

Solutions are required to maintain methods of long term clinical trial follow-up that accurately capture outcome data. Routinely collected data from hospitals linked to cancer registries may provide some solutions for long term clinical trial follow-up, without the need for additional clinic visits as demonstrated by Appleyard et al in this Special Issue (). We will build on this concept and discuss preliminary work investigating the application and feasibility of using routine data sources for the purposes of identifying accuracy of baseline data, recurrence and survival data for cancer clinical trial follow-up.

What is the feasibility of obtaining cancer clinical trial follow-up from routine data sources?

The National Cancer Registration and Analysis Service (NCRAS) are responsible for the systematic collection, quality assurance and analysis of cancer registration data in England to support cancer epidemiology, public health, service monitoring and research. The NCRAS routine datasets may provide valuable information regarding baseline patient demographics and characteristics, treatments, safety, survival and recurrence data (3).

Although routine data sources were designed to collate cancer outcome data for patients, they were not specifically designed to collect information for use in clinical trials and the focus of the NCRAS data is not to support long-term clinical trial follow-up. Clinical trial related endpoint data may not be easily available in routine data. For example, local recurrence is an important endpoint in cancer clinical trials, but the Cancer Outcomes and Services Dataset (COSD) does not differentiate between local and distant recurrences. In addition, major cardiovascular events may be identifiable, but radiotherapy related normal tissue effects (NTEs) would not be available. Therefore although NCRAS may provide a vast amount of data, the *feasibility* of using routine data sets for long-term trial follow-up needs assessment. Preliminary

work investigating whether routine data can be used for long-term trial follow-up is ongoing with several groups.

TACT- NCDR Retrospective linkage project

The Institute of Cancer Research Clinical Trials and Statistics Unit (ICR-CTSU) and NCRAS have conducted a joint retrospective linkage project comparing routine data in the National Cancer Data Repository (NCDR) with the TACT trial *(ISRCTN:79718493)* data (4). The aim of this study was to assess quality and completeness of the routine data and to identify the current ability of routine datasets to determine cancer trial outcomes as a baseline to determine the feasibility of a prospective validation study. Major findings from the TACT-NCDR linkage project included (i) lack of standardisation of data collection across registries requiring extensive data cleaning and (ii) a large amount of missing data. Of note, patients were included in the TACT study over 10 years ago and therefore we would expect the registry data available now to be improved with less missing data. A third finding was the lack of recurrence data and/or the inability to determine whether this was local or systemic. Preliminary work by NCRAS on the development of an algorithm to characterise recurrences using routinely collected data has now been done. This algorithm requires further refinement and validation which will be conducted by the ICR-CTSU in partnership with NCRAS and will now be outlined.

NCRAS Prospective validation study

Following on from the TACT-NCDR linkage study, the next step is to prospectively evaluate the new routine datasets; COSD, Systemic Anti-Cancer Therapy Dataset (SACT) and National Radiotherapy Dataset (RTDS) with contemporary trial data. This analysis is timely as these datasets are now available for analysis and there is an agreement in place for the ICR-CTSU to work together with NCRAS on this study. The NCRAS prospective validation study aims to identify whether routine datasets

are of sufficient standard to replace traditional data collection methods and also to help NCRAS improve their data collection and identify areas of quality concern where data collection can be improved. The ICR-CTSU will work with NCRAS to prospectively validate COSD, SACT, RTDS and Hospital Episode Statistics (HES) datasets using data from four early breast cancer trials TACT2 (ISRCTN68068041)(5), POETIC (ISRCTN63882543), IMPORT HIGH (ISRCTN47437448) and FAST FORWARD (ISRCTN19906132) deliberately chosen to provide a range of time periods of recruitment and treatment modalities. Data will be requested and analysed over a period of 5 years in the first instance. (See table)

The objectives of the prospective validation study include identifying and quantifying trial participants within each dataset and assessing the completeness, validity and consistency of routine data with trial data. In particular, time to tumour recurrence, overall, disease-free and relapse-free survival, time to loco-regional and distant recurrence and long term safety/ comorbidity data in TACT2 and POETIC and local control, relapse and disease free survival, time to distant metastases and late adverse effects in IMPORT HIGH and FAST FORWARD. A cross comparison of trial baseline and treatment data and emerging disease related outcome data with routine data will also be undertaken. Long-term safety data will be collected and the representativeness of trial patients versus the general population identified. Once the project is fully established the plan is to extend beyond breast cancer into other disease sites including bladder and prostate cancers.

Prospective validation of routine data in recruiting studies

In addition to prospectively validating routine cancer registry data with trial data from completed studies, mechanisms for collecting data directly from cancer registry data in recruiting studies will be established within the ICR-CTSU. In the PRIMETIME *(ISRCTN: 41579286)* (6) 'Post-operative avoidance of radiotherapy: biomarker selection of women categorised to be in a very low risk group by IHC4+C' in early

breast cancer study, data (including local, distant relapse, new primary cancers, death, cause of death) will be collected from routine data sources. These routine data sources will include COSD, RTDS, SACT, HES, the Office for National Statistics (ONS) mortality file and the equivalent databases in the devolved nations. These datasets will be used to compare routine data collection with the traditional data collection methods to identify ipsilateral breast local relapse. Six monthly data downloads will be received from NCRAS. One of the main objectives in PRIMETIME is to obtain routine data regularly and with a quick turnaround to identify if we are able to detect events from the routine data sources in 'real time'. If routine data is used in place of traditional data collection methods the routine data needs to be frequently updated and any lag periods avoided. The time spent requesting, receiving and processing routine data will need to be reduced as much as possible to avoid delays in detecting an outcome which could compromise trial reporting.

Conclusion

Routine NHS cancer data may provide a source of baseline, recurrence, survival and toxicity data for clinical trial follow-up and the feasibility of this application is under investigation. The TACT-NCDR retrospective linkage study demonstrated the need for standardisation of data collection and extensive data cleaning. Further improvements may be demonstrated following an increase in cancer registry data quality and development of an algorithm to characterise recurrences. This will be assessed within the ongoing NCRAS prospective validation study and will give insight as to whether routine datasets are of sufficient standard to replace traditional data collection methods. In addition, validation and refinement of the algorithm to characterise recurrences may allow discrimination between local and distant recurrences from routine data sources. Future challenges such as collection of long term toxicity data are ongoing, but by working in partnership with Public Health England and NCRAS we hope to identify which routine data sources are a viable

source of long-term follow-up clinical trial data and develop ways of improving data quality and the patient follow-up process.

Table 1
Summary of trials and data in the National Cancer Registration and Analysis Service (NCRAS) prospective validation study

Trial overview	Trial data to be investigated in the NCRAS prospective validation study
TACT2: multicentre phase III trial examining optimal chemotherapy regimen administration following surgery. Participants were randomised into one of four treatment groups (E-CMF; aE-CMF; E-X; aE-X) in a 1:1:1:1 ratio.	Time to tumour recurrence Overall survival Disease-free survival Relapse-free survival Time to distant recurrence Time to locoregional recurrence Long-term safety/incidence of comorbidities
POETIC: multicentre phase III trial testing perioperative aromatase inhibitor therapy followed by standard adjuvant therapy in postmenopausal women compared with standard adjuvant therapy alone. Patients allocated in a 2:1 ratio to perioperative therapy with an aromatase inhibitor for 4 weeks (2 weeks before and 2 weeks after surgery) or no perioperative therapy.	Time to tumour recurrence Relapse-free survival Time to local recurrence Time to distant recurrence Overall survival Breast cancer-free survival
IMPORT HIGH: multicentre, randomised phase III trial testing dose-escalated intensity-modulated radiotherapy for women treated by breast-conservation surgery and appropriate systemic therapy for early breast cancer. Participants randomised into one of three treatment groups (sequential boost dose: 56 Gy/23 fractions, concomitant boost dose: 48 Gy/15 fractions, concomitant boost dose 53 Gy/15 fractions) in a 1:1:1 ratio.	Local tumour control Contralateral primary tumours Other primary tumours Regional metastases-free survival Distant metastases-free survival Overall survival Late adverse effects
FAST Forward: multicentre, randomised phase III trial testing a 1 week course of whole breast radiotherapy against a standard 3 week schedule in terms of local cancer control and late adverse effects in patients with early breast cancer. Participants were randomised into one of three treatment groups (control: 40 Gy/15 fractions, test group 1: 27 Gy/5 fractions, test group 2: 26 Gy/5 fractions) in a 1:1:1 ratio.	Ipsilateral local tumour control Relapse-free survival, Disease-free survival, Time to distant metastases Overall survival Late adverse effects

References

1. Kumar HS. National Health Service Cancer Strategy for England: Where Are We Now? Clinical oncology (Royal College of Radiologists (Great Britain)). 2016;28(5):277-8.
2. Appleyard S, Gilbert DC. Innovative solutions for clinical trial follow-up: adding value from nationally held UK data. Clin Oncol 2017;29:789e795.
3. Victora CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. American journal of public health. 2004;94(3):400-5.
4. Kilburn L AM, Banerji J and Bliss J. Results of a retrospective linkage of TACT (CRUK 01/001) breast cancer trial data and the National Cancer Data Repository. National Cancer Research Institute (NCRI) Cancer Conference 2012; 4-7 Nov; Liverpool, UK. NCRI; Nov 2012. Abstract A141. Available from: http://abstracts.ncri.org.uk/abstract/results-of-a-retrospective-linkage-of-tact-cruk-01001-breast-cancer-trial-data-and-the-national-cancer-data-repository-2/.
5. Cameron D, Morden JP, Canney P, Velikova G, Coleman R, Bartlett J, et al. Accelerated versus standard epirubicin followed by cyclophosphamide, methotrexate, and fluorouracil or capecitabine as adjuvant therapy for breast cancer in the randomised UK TACT2 trial (CRUK/05/19): a multicentre, phase 3, open-label, randomised, controlled trial. The Lancet Oncology. 2017.

6. Kirwan CC, Coles CE, Bliss J. It's PRIMETIME. Postoperative Avoidance of Radiotherapy: Biomarker Selection of Women at Very Low Risk of Local Recurrence. Clinical oncology (Royal College of Radiologists (Great Britain)). 2016;28(9):594-6.