A comparison of the effectiveness of different treatment regimens for pancreatic cancer using English cancer registry data

Background

Large amounts of data are collected on cancer patients in the National Health Service (NHS), held by the National Cancer Registration and Analysis Service (NCRAS). The Systemic Anti-Cancer Therapy (SACT) database, part of the NCRAS dataset, collects information on systemic-anti cancer therapies on a national scale and can be linked to other data sources (such as hospital episode statistics (HES) and radiotherapy datasets) to provide a complete picture of the cancer patient pathway. NCRAS has been commissioned by NHS England (NHSE) to provide data and analysis for the evaluation of drugs that are in the Cancer Drugs Fund (CDF), with the aim of using the data to resolve uncertainties around the effectiveness and cost-effectiveness of cancer treatments placed in the CDF. Despite this, as yet, no attempts have been made to assess whether the data held by NCRAS is sufficient for reliably comparing the effectiveness of different cancer treatments.

Why this research is needed now

Estimating the comparative effectiveness of treatments using observational data (such as NCRAS data) is prone to important biases, because the treatment that people receive may be strongly related to their prognostic characteristics, creating a selection bias. Time-dependent confounding can also be an important issue, if people change treatments over time. This is why randomised controlled trials (RCTs) are usually used to compare the effectiveness of different treatments. However, given that lots of data are collected on cancer patients in the NHS, and that these data are being used to inform treatment recommendations, it is important to investigate whether analyses of these data can provide reliable results.

Aims and Objectives

We will use NCRAS data to emulate four existing RCTs of treatments for pancreatic cancer. We will attempt to replicate the eligibility criteria and treatment strategies specified in the RCTs as closely as possible, and will use causal inference statistical methods to attempt to avoid bias. We will compare the outcomes of our analyses to the results of the existing RCTs, as a way of determining whether our NCRAS-based analyses produce reliable estimates of treatment effectiveness. In addition, for each trial emulation, we will also undertake analyses for the broader NHS population (i.e. not restricted to the population defined by the RCT eligibility critieria).

The RCTs that we will attempt to emulate are:
- ESPAC-4. Comparing gemcitabine monotherapy with gemcitabine plus capecitabine in patients with adjuvant pancreatic cancer
- ACCORD. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer
- CRUK-GEM-CAP. Gemcitabine versus gemcitabine plus capecitabine for metastatic pancreatic cancer
  MPACT. Gemcitabine versus gemcitabine plus nab-paclitaxel for metastatic pancreatic cancer

Funding

Protecting Patients

This study involves analysing NCRAS data, including cancer registration datasets, the SACT dataset, the radiotherapy dataset, Hospital Episodes Statistics data, and the cancer waiting times dataset, for patients who received systemic anti-cancer therapy for pancreatic cancer between 2012 and 2021. This is necessary to allow us to identify patients to include in our RCT emulations, and to run statistical analyses that control for prognostic factors that could cause bias.

Patient identifying information already held by the NHS will be used by Public Health England (PHE) to link the different datasets. This will be done entirely by PHE. No patient identifying information (such as name, date of birth or address) will be made available to the researchers. All identifiable data will be removed by the PHE data analysts before any data is made available to the research team; all data received by the research team will be 'pseudonymised' (i.e. PHE will assign a non-identifying ID number to each record, in place of the deleted identifiable information such as names, dates of birth etc.). We also obtained the approval of both the NHS Research Ethics Service before the release of any data. No record-level data used in this study will be transferred to any third country or international organisation at any point. The overall Data Retention Period for the project ends on 4/02/2027, and all original personal data will be securely deleted by this date.

Since the data have been pseudo-anonymised and are part of a large routine data set, the identities of individuals are unknown to the researchers and there is no way of opting out of the data analysis. However, the data were supplied by the NHS which implements the national data opt out programme. For further details, please see: https://www.nhs.uk/your-nhs-data-matters/

Data Controller

The University of Sheffield is the **Data Controller** for this study. Please see the University of Sheffield **General Privacy Notice** for information on the handling of personal data by the university for research (e.g. this study), including details of the data protection officer, supervisory authority, right of complaint, and lawful basis for data processing under GDPR.

Contact Information

Professor Nicholas Latimer
School of Health and Related Research.
Regent Court,
30 Regent Street,
S1 4DA Sheffield, UK
n.latimer@sheffield.ac.uk