CASNET 2 Statistical Analysis Plan

Outcomes:

1. Primary care interval for cancer diagnosis
    a. Compare active with inactive periods
2. Proportion of cancers diagnosed after emergency presentation
    a. Compare active with inactive periods
3. Recorded new diagnoses in instances of ESN toolkit activation
4. Total time to diagnosis
    a. Compare active with inactive periods
    b. All cancer diagnoses and all diagnoses with template activation
5. Number of primary care consultations between 1st recorded symptom and referral
    a. Compare active with inactive periods
6. Rates of patients completing direct access cancer investigations
    a. Compare active with inactive periods
7. Referral rates, by urgency route (2 week wait, urgent, routine)
    a. Compare active with inactive periods
8. Timing of template activation within the primary care interval
9. Template activation rate amongst consulting patients
    a. Total and stratified by GP
10. Proportion of diary entries that were completed
11. Reason for template activation (based on 20 high level READ codes)
12. Symptoms leading to direct access to investigations
13. Recorded vague symptoms in the template
14. Demographic details of patients with activated templates
15. GP type completing templates (e.g. partner, locum, trainee)

PPI prioritised outcomes

These will be analysed as a priority.

- Primary care interval for cancer diagnosis (primary outcome)
- Total time to diagnosis
- Number of primary care consultations between 1st recorded symptom and referral
- Rates of patients completing direct access cancer investigations
- Proportion of cancers diagnosed after emergency presentation
- Recorded new diagnoses in instances of ESN toolkit activation
- Referral rates, by urgency route (2 week wait, urgent, routine)

Data inclusion

Analyses will include all participants. Practices that withdraw their agreement to participate will be included in analyses up to the point of withdrawal, unless they indicate that they wish to withdraw previously collected data from analysis.

The date of template switch-on will act as the indicator of becoming ‘post’ – acts as an index date. There may be some bias in this approach as patients who consult prior to switch-on date, but whose

referral is made after this date will be included in the post-intervention period. This is likely to bias towards the null as they were not actually exposed to the intervention.

Patients will be classified into "pre" or "post" on the basis of the date of their cancer referral, or diagnosis if earlier.

Study time periods

Start date	End date	Period	Relevant variables
12-Oct-2018	11-Oct-2019	Before study period	cancer symptoms only
12-Oct-2019	11-Oct-2020	Pre-intervention	All, including symptoms and diagnoses
12-Oct-2020	19-Aug-2021	Switch-on	All, including symptoms and diagnoses
20-Aug-2021	30-Nov-2021	Post-intervention	All, including symptoms and diagnoses
1 Dec-2021	31-May-2022	Follow-up	Cancer diagnosis only

Definition and timing of variable measures

In the analyses, a year is defined as 365 days.

Date of cancer diagnosis (cancer patients): after the start date of the study and within the data collection period. Where patients have multiple cancer diagnoses, we will choose the first cancer diagnosis to define this. We will report the number of people with multiple cancer diagnoses, and, if possible, summarise this information including number of diagnoses, time between diagnoses, and body systems. We will also summarise the number of cancers diagnosed in the before study period, with subsequent diagnoses in the study period.

Date of cancer referral (cancer patients): after the start date of the study. For primary analysis, we will use the closest referral preceding the diagnosis. As it's possible that this could identify referrals that do not relate to the cancer, sensitivity analysis will be carried out by using the closest 2ww referral (if present) preceding the diagnosis, and matching the site of the diagnosis.

Date of first cancer symptom (cancer patients): first cancer symptom in the year prior to the diagnosis. We will tabulate the frequency of each first symptom by cancer site.

Date of emergency department presentation (cancer patients): after the start date of the study. Closest presentation preceding the diagnosis.

Date of toolkit activation (ESN patients): first toolkit activation after the start of the intervention.

Date of toolkit activation (pre activations): first toolkit activation before the start of the intervention.

ESN cancer diagnoses (ESN patients): all cancer diagnoses after the ESN activation

ESN non-cancer diagnoses by body system (ESN patients): all non-cancer diagnoses after the ESN activation for patients without a cancer diagnosis

Primary care consultation (all patients): a date on which one or more events occur which are coded as either Home Visit, Face to Face, Video, Telephone, or Unclear Consultation Mode using the Laura-and-Brian coding method.

Number of primary care consultations in primary care interval: number of primary care consultation in the time between 1st recorded symptom and referral, including the dates of symptom and referral.

Number of patient years before intervention (all patients): total patient years between max(date of registration, date of study start) and min(date of intervention, date of death, date of deregistration)

Number of patient years during intervention (all patients): total patient years between max(date of registration, date of intervention) and min(date of study end, date of death, date of deregistration)

Dates of all direct access cancer investigation: date of every direct access investigation

Referral type (2ww/urgent/routine) for cancer referrals (cancer patients): referral type for the closest referral before a cancer diagnosis

Referral type (2ww/routine) for all referrals: referral type for all referrals in the study period

Number of patients with a consultation before activation: total number of patients with at least one primary care consultation between the start of the study and start of intervention

Number of patients with a consultation after activation: total number of patients with at least one primary care consultation between the start of intervention and the end of the study

Coded reasons for template activation: template reason codes recorded on the same day as a template activation

Cancer symptoms on the same day as referral to direct access investigation: all cancer symptoms coded on the same day as a referral to direct access investigation

Dates of all toolkit activations (ESN patients): all toolkit activation after the start of the intervention.

Cancer symptoms recorded on the same day as ESN activation: all cancer symptoms coded on the same day as any template activation

Age: age at time of study start

Sex: sex at study start

Ethnicity: coded ethnicity at study start

IMD quintile: IMD quintile at study start

Definition of Outcome Measures

Primary care interval for cancer diagnoses

The primary care interval is defined as the number of days between 1st recorded symptoms of cancer (within the year prior to diagnosis) and subsequent referral for secondary cancer care.

Population: all patients with a cancer diagnosis

Underlying variables: Date of cancer diagnosis, date of cancer referral, date of first recorded cancer symptom in the year before diagnosis

Analysis: compare pre and post activation groups using mixed effects linear regression, adjusting for defined covariates.

Imputation:

- Where the symptom occurs after the referral, or there is no symptom:
  - Report numbers affected
  - Impute at 1 year
- Where date of referral is after date of diagnosis
  - Use time to referral, even if this is longer than time to diagnosis
- Where there is no referral date
  - Report numbers affected
  - Impute at one year

Sensitivity analysis:

- excluding imputed primary care intervals
- defining referral by closest 2ww referral

Proportion of cancers diagnosed after an emergency presentation

The proportion of total cancer diagnoses where the diagnosis is made soon after an emergency department presentation.

The plan in the protocol was: "We will develop algorithms to identify emergency presentations of cancer (where a diagnosis of cancer is made prior to a referral), including following an attendance at A&E or an inpatient episode (35). Where there is uncertainty regarding the route of diagnosis, the PLO team within the RCGP RSC network will contact the practice in an attempt to augment the data. " However, delays from COVID means we are no longer able to check via the PLO.

Following the "broad definition" defined in the Supplemental data for McPhail et al (https://doi.org/10.1016/S1470-2045(22)00127-9), we will therefore define an emergency presentation of cancer as a presentation at an emergency department within the 30 days preceding the cancer diagnosis (irrespective of the date of referral.)

Population: all patients with a cancer diagnosis

Underlying variables: Date of cancer diagnosis, date of emergency department presentation

Analysis: compare pre and post activation groups using mixed effects logistic regression, adjusting for covariates.

Recorded new diagnoses in instances of E-SN toolkit activation

Coded entries for all alternative diagnoses where the E-SN toolkit has been activated (cancer and non-cancer).

Population: all ESN patients

Underlying variables: Date of ESN toolkit activation, ESN cancer diagnoses, ESN non-cancer diagnoses by body system.

Analysis: descriptive

- Proportion of patients with cancer
- Breakdown of cancers by body system
- Breakdown of non-cancers by body system (in non-cancer patients only)

Total time to diagnosis

1st recorded symptom of cancer (within the previous year) to definitive diagnosis for all cancers diagnosed, and for all patients with an activated template.

Population: all patients with a cancer diagnosis

Underlying variables: Date of cancer diagnosis, date of first recorded cancer symptom in the year before diagnosis

Analysis: compare pre and post activation groups using mixed effects linear regression, adjusting for defined covariates.

Secondary population: all patients with an ESN activation before the cancer diagnosis

Underlying variables: Date of cancer diagnosis, date of first recorded cancer symptom in the year before diagnosis

Analysis: compare pre (all cancer patients) and post (cancer + ESN) activation groups using mixed effects linear regression, adjusting defined covariates – per protocol analysis

Imputation:

- Where the symptom occurs after the diagnosis, or there is no symptom:
    - Report numbers affected
    - Impute at 1 year

Sensitivity analysis:

- excluding imputed primary care intervals

Number of primary care consultations between 1st recorded symptom and referral

Number of primary care consultations between the 1st recorded symptoms (within the year prior to diagnosis) and subsequent referral, per patient.

Population: all cancer patients

Underlying variables: number of primary care consultations in primary care interval.

Analysis: compare pre and post activation groups using mixed effects Poisson regression, adjusting for defined covariates

Imputation:

- Where the symptom occurs after the referral, or there is no symptom:
    - Impute at 1 year
- Where date of referral is after date of diagnosis
    - Use time to referral, even if this is longer than time to diagnosis
- Where there is no referral date
    - Impute at one year

Sensitivity analysis:

- excluding imputed primary care intervals
- defining referral by closest 2ww referral

Rates of patients completing direct access cancer investigations

The numerator will be the number of patients undergoing direct access cancer investigations (according to those specified in referral guidelines NG12 (36, 37)) in each period divided by the person years of observation for that period.

Population: all patients

Underlying variables: Number of patient years before intervention, number of patient years during intervention, dates of all direct access cancer investigations

Include all direct access investigations, even if there are multiple per person

Analysis: compare pre and post activation groups using mixed effects Poisson regression, adjusting for defined covariates, and with random effect for patient

Referral rates, by urgency route (2 week wait, urgent, routine)

For all patients referred for specialist opinion to a secondary care cancer specialist, information will be ascertained about the route of referral (two-week wait, urgent, routine).

Population: all cancer patients

Underlying variables: referral type (2ww/urgent/routine) for cancer referrals

Secondary population: all patients with a referral

Underlying variables: referral type (2ww/routine) for all referrals

Analysis: logistic mixed regression for each referral type, with random variable by patient for secondary pop

Timing of template activation within the primary care interval

The number of days between 1st recorded symptoms (within the year prior to diagnosis) and template activation and the number of days between template activation and subsequent referral.

Population: all ESN patients with cancer diagnosis

Underlying variables: Date of template activation, date of cancer referral, date of referral

Exclude if ESN is outside primary care interval

Analysis: descriptive

- Min, max, median, IQR
- Histograms

Template activation rate amongst consulting patients

The number of patients with an activated template divided by the number of patients consulting (Total and stratified by GP practice)

Population: all patients

Underlying variables: Number of patients with a consultation before activation, number of patients with a consultation after activation, Date of toolkit activation (ESN patients and pre-activations), Practice ID for template activations

Analysis: descriptive

- Total rate before activation (hopefully zero)?
- Total rate after activation
- Rate by practice

Subgroup analyses

- By cancer diagnosis

Proportion of diary entries that were completed

Planned: The number of diary entries that were completed divided by the number of diary entries that were opened.

Due to COVID pressures limiting data acquisition of the practice downloads, it is not possible to assess this outcome.

Reason for template activation (based on 20 high level READ codes)

The coded reasons for activating the template.

Population: all patients with an ESN activation (could be multiple per patient)

Underlying variables: coded reasons for template activation

Analysis: descriptive

- Proportion of each code

Symptoms leading to direct access to investigations

Recorded symptoms prior to direct access investigations – on the day of referral to direct access

Population: all patients receiving direct access investigations

Underlying variables: Dates of all direct access investigation, cancer symptoms on the same day as referral to direct access investigation

Analysis: descriptive

- Proportions of each symptom
- By type of investigation

Subgroup analyses:

- By template activation status (before the direct access investigation)
- By having a cancer diagnosis (after investigation)

Recorded vague symptoms in the template

All symptoms recorded within the template. (on the same day as template activity)

Population: all patients with an ESN

Underlying variables: Date of all ESN activations, vague symptoms recorded on the same day as ESN activation.

Analysis: descriptive

- Proportions of each symptom

Subgroup analysis:

- By cancer diagnosis (after template activation)

Demographic details of patients with activated templates

Age and sex of patients that had a template activated during the course of the trial.

Post hoc analysis: ethnicity and IMD quintile of patients that had a template activated during the course of the trial.

Population: all patients with an ESN

Underlying variables: Age at activation, recorded sex, recorded ethnicity, recorded IMD quintile

Analysis: descriptive

- Proportions of each sex, ethnicity, and IMD quintile
- Mean (SD) of age
- Histogram of age

Subgroup analysis:

- By cancer diagnosis (after template activation)

GP type completing templates (e.g. partner, locum, trainee)

Planned: Descriptive data on the type of GP that first activated the template

This analysis will not be carried out as it is not possible to identify GP types in the data, and it is not appropriate to look at individual GPs due to the risk of small numbers in the outcomes.

Post hoc analysis: Number of ESN activations per person

Summary of the number of ESN activations by person.

Population: all patients with an ESN

Underlying variables: Dates of all ESN activations

Analysis: descriptive

- Min, max, median and IQR of numbers
- Histogram of numbers

Subgroup analysis:

- By cancer diagnosis (after template activation)

Analysis Methods

Descriptive outcomes

The following outcomes are descriptive only:

- Recorded new diagnoses in instances of ESN toolkit activation
- Timing of template activation within the primary care interval
- Template activation rate amongst consulting patients
- Reason for template activation (based on 20 high level READ codes)
- Symptoms leading to direct access to investigations
- Recorded vague symptoms in the template
- Demographic details of patients with activated templates

Details of exact descriptive statistics are given above, but may include rates, proportions, and mean/SD or median/IQR as appropriate

Comparative outcomes (active vs inactive periods) using multi-level regression

- Primary care interval for cancer diagnosis (linear regression)
- Proportion of cancers diagnosed after emergency presentation (Logistic regression
- Total time to diagnosis (linear regression)
- Number of primary care consultations between 1st recorded symptom and referral (poisson regression)
- Rates of patients completing direct access cancer investigations (Poisson regression, with random effect for participant)
- Referral rates, by urgency route (Logistic regression, with random effect for participant in the case of non-cancer diagnoses)

For linear regression of times, histograms of the outcome will be viewed to determine whether logarithmic transformation is needed. If transformation is deemed appropriate, sensitivity analysis without logarithmic transformation of the outcome will be carried out.

These outcomes will be analysed using hierarchical (mixed models) regression, using a fixed effect for the intervention condition, a fixed effect for age, sex, and ethnicity at the patient level, a fixed effect for urban/rural status and the time of activation at the practice level, a random effect for practice, and where necessary, a random effect for participant. Initial analysis will determine whether IMD has been calculated at the patient or practice level, and it will be included as a fixed effect at the relevant level.

Sensitivity analyses:

- excluding patients in the post group who started the diagnostic journey before the introduction of the E-SN toolkit
- without logarithmic transform of linear time outcomes, if logarithmic transformation is used.

Planned subgroups

Where applicable (see outcome definitions), subgroups will be:

Planned subgroup analyses will be carried out for the following outcomes

Patients in whom an E-SN toolkit entry was completed

- Per-protocol analysis for all comparative outcomes, where we compare only ESN patients in the post group to all patients in the pre group
- Symptoms leading to direct access to investigations
- Rates of patients completing direct access cancer investigations
- Referral rates, by urgency route

Patients diagnosed with cancer

- Template activation rate amongst consulting patients
- Demographic details of patients with activated templates
- Reason for template activation (based on 20 high level READ codes)
- Recorded vague symptoms in the template
- Symptoms leading to direct access to investigations (comparing only cancer patients in both time periods)
- Rates of patients completing direct access cancer investigations (comparing only cancer patients in both time periods)
- Referral rates, by urgency route (comparing only cancer patients in both time periods)

Planned sensitivity analysis

Excluding patients overlapping groups

For all comparative outcomes, we will undertake a sensitivity analysis excluding patients in the post group who started the diagnostic journey before the introduction of the E-SN toolkit to estimate the effect once the E-SN toolkit is universally available. This will be patients whose first symptom is in the pre-intervention period, but whose referral (or diagnosis if earlier) is in the post-intervention period.

Excluding imputed primary care intervals

Analyses excluding imputed primary care intervals will be carried out for:

- primary care interval for cancer diagnosis (primary analysis)
- Total time to diagnosis
- Number of primary care consultations between 1st recorded symptom and referral

Defining referral by closest 2ww referral

Analyses using cancer referral defined by the closest 2ww referral in the matching body system will be carried out for:

- primary care interval for cancer diagnosis (primary analysis)
- Number of primary care consultations between 1st recorded symptom and referral

Post hoc analyses

Time interval from referral to diagnosis

The time intervals from referral to diagnosis will be summarised in both groups. A comparative analysis using multi-level regression will be carried out for this time interval, similar to the primary analyses but excluding sensitivity and subgroup analyses.

Data display and reporting

We will combine or suppress any cells with small numbers (under 5) of observations to prevent any potential identification during the reporting of the results.