Public Health England

Protecting and improving the nation's health

Travel times and cancer

Learning from pilots and standardising methodology

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing, and reduce health inequalities. We do this through world-leading science, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. We are an executive agency of the Department of Health and Social Care, and a distinct delivery organisation with operational autonomy. We provide government, local government, the NHS, Parliament, industry and the public with evidence-based professional, scientific and delivery expertise and support.

Public Health England
Wellington House
133-155 Waterloo Road
London SE1 8UG
Tel: 020 7654 8000
www.gov.uk/phe
Twitter: @PHE_uk
Facebook: www.facebook.com/PublicHealthEngland

© Crown copyright 2018
You may re-use this information (excluding logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0. To view this licence, visit OGL. Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

Published June 2018
PHE publications
gateway number: 2018190

Corporate member of
Plain English Campaign
Committed to clearer
communication
339

PHE supports the UN
Sustainable Development Goals

Contents

Key messages

- Travel times to all major cancer service providers have been pre-calculated, allowing data analysis by travel time without individual access to specialist software.
- Identifying and defining cancer centres and specialist service providers remains a key challenge for travel time analysis.
- Minor variations in the methodology used to calculate travel times exist between current publications. While moving towards a standard methodology, we acknowledge that there are advantages and disadvantages of any methodological choice, and specific questions may require tailored methodology.

Background

In 2015 the Independent Cancer Taskforce recommended an evaluation of the impact on cancer outcomes of patients living different distances from a cancer centre (recommendation 27)[1]. With evidence showing that centralisation of services can improve outcomes for certain treatments, it is important to balance centralisation with the implications for patients of having to travel further.

In order to address this recommendation the National Cancer Registration and Analysis Service (NCRAS) has investigated different ways of calculating travel times. Pilot work has included a study comparing survival rates by travel time to nearest hospital with a multidisciplinary team (MDT) for patients with breast, prostate, lung and colorectal cancer; and a site specific study on prostate cancer investigating whether travel time to a radiotherapy centre appears to affect choice of radiotherapy as a treatment.

This document sets out the principles that should be considered when producing travel times analysis, based on the methodological issues that these early projects have highlighted.

he methodological issues identified have been grouped into 4 areas:

- identifying start and end points of patient journeys
- calculating the time the journeys are expected to take
- making the calculated travel times available for all analysts
- presenting travel times

Identifying start and end points of patient journeys

Embarkation: Start points of journeys

The NCRAS collects data on the patient's home postcode at diagnosis. In previous travel time analysis, this postcode has been assumed to be the start point of the patient's journey. This data item is highly complete, with a postcode at diagnosis recorded for over 99.9% of patients.

There may be limitations to using this field. Patients may move during their care for a variety of reasons, which may be related to their cancer, for example: to be nearer to their treatment, nearer to the support of their family, or to access residential care or nursing care. The postcode at diagnosis is less likely to be accurate the more time has passed since diagnosis, so this may be a particular problem for analysis of late patient pathways and survivorship, for example: more than 5% of patients no longer have the same home postcode on their hospital records within one year of their cancer diagnosis, rising to over 10% after 5 years. The full length of the journey for treatment may not just be from the patient's home to the hospital, if for example a relative is providing a lift.

The cohort of patients included will need to be defined geographically. Most NCRAS analysis includes patients resident in England when their cancer is diagnosed. This includes patients resident on islands which are not connected by the road network to the mainland, mainly the Isle of Wight (the largest in terms of both size and population) and the Isles of Scilly (the Isle of Man and the Channel Islands are not part of England, they are self-governing Crown Dependencies, and are usually excluded from analysis). Although there are limited hospital services in both locations (a small NHS Trust which has no inpatient oncology on the Isle of Wight, and a GP led community hospital on the largest of the Isles of Scilly) patients have to travel to the mainland for the majority of their cancer treatment. These long travel times are outliers in the cohort of all English patients. It is recommended that when analysing trends by travel time sensitivity analysis is performed including and excluding patients on these islands.

Destination: end points of journeys

Identifying an end point of the patient journey can be divided into 3 levels of increasing methodological difficulty:

1 The cohort of patients all have the event of interest, and the event happens in a unique known location for each patient. For example, with the question 'how long do patients receiving stereotactic radiotherapy have to travel for?', each patient is treated in one specialist centre for stereotactic radiotherapy, the location of their treatment is recorded by NCRAS, and the journey can be calculated between their postcode at diagnosis and their treatment location.

2 The cohort of patients all have the event of interest, but the event may not happen in a unique location for each patient. This situation can arise, for example, with the question 'how far do patients have to travel for their cancer treatment?' The patient pathway is complex for many cancer patients, with diagnosis, surgery, chemotherapy and radiotherapy often taking place in different locations. To calculate the travel times for 'treatment', a precise definition of the treatment(s) of interest, and a methodology for defining the travel of interest for patients with treatment in multiple locations is needed. The best solutions to this often depend on the precise question being asked. Standard NCRAS algorithms now exist to determine some locations, such as 'hospital of diagnosis'.

3 The cohort of patients do not all have the event of interest, and the hypothesis under investigation is whether travel time has affected this. For example 'do patients living near a lung cancer specialist centre get more lung cancer surgery?', or 'are patients living nearer to a radiotherapy centre more likely to choose radiotherapy as a treatment for prostate cancer?' In this case, patients who do not have the event do not have a recorded location, and their 'nearest centre' or 'catchment centre' needs to be calculated. This relies on having a full list of locations where the event happens, which may be challenging to produce (see next section). Generally the pilot studies have calculated the 'nearest' hospital which may not fully represent patient flows and catchment areas. In order to avoid bias, the 'nearest centre' is usually calculated for all patients in the cohort of interest, even those that have had the event of interest (which may not have happened in their nearest centre).

Specifying the location of services

Many travel times projects are investigating questions of the form of (3) above - the travel time to the nearest place of interest. The place of interest will depend on the question being asked and the expected patient pathway - for example a cancer centre, a site specific specialist centre, a radiotherapy centre or a doctor's surgery.

To do this, a list of *all* these centres is needed, so that the nearest can be found. Initially, analysts aimed to develop a central comprehensive list of places of interest that could be used for all travel time calculations. This may be possible for a defined set of commonly asked questions, but maintaining such a list is a complex project that varies over time (with hospitals changing their physical location as new buildings are constructed, and with specialised services moving between hospitals). It is also very cancer site dependent as the hospitals that have an MDT are different for different cancer sites, for example the pilot study on survival worked with 4 different lists, one for each cancer site.

Care should be taken to consider likely patient flows near to borders. For highly specialised cancer services, there may be service providers in Wales or Scotland that are nearer than the closest English service for patients living very close to the border. Talking to local specialists to understand the cross-border arrangements in these areas should resolve whether these service providers should be included on the lists of 'all centres'.

The NCRAS recommends that analysts may need to develop project specific lists of centres. These should be developed by:

a) reviewing other lists used previously for travel time analysis
b) reviewing publically available documentation on where care is provided, such as the Quality Surveillance Team lists
c) producing data-driven lists created by querying registry data to find out where the events of interest are recorded to have happened
d) resolving conflicts between (a), (b) and (c) by talking to local specialists, consulting the NHS Choices web site2, individual Trust websites, and talking to service providers (hospitals) directly

The list of centres should be published with the analysis. As more analysis is done in this area, a more comprehensive library of lists of places of interest will be produced over time (by year, cancer site, treatment type etc).

Quality of data on location of cancer services

Once the end point of the journey has been identified, a precise postcode is needed to calculate the travel time.

The service provider is often recorded at Trust level instead of hospital level. This data is sufficient to calculate travel times if all of the Trust is based on one campus location. However, many Trusts have more than one physical site which may have a large effect on travel times - for example York Teaching Hospital NHS Foundation Trust includes hospitals that are 60km apart.

We are working with the data liaison teams to encourage data submission with full hospital codes. For diagnosis events which took place in in 2015, the cancer registry recorded 59% of them with a full hospital code, while 39% had only a Trust code, and 2% had no location recorded.

A 'Trust diameter' has been calculated for each Trust code. If the Trust diameter is 0, then the Trust is all on one site, and so can be used as an accurate endpoint for travel time analysis despite no hospital code being recorded. The wider the diameter of the Trust, the greater the potential for error in travel time calculations. 50% of Trusts are all on one campus, and a further 10% have diameters of less than 5km. There are 10 trusts with a diameter of more than 35 km.

We have also produced a list of 'campuses'. Each entry on this list is one physical location where cancer care is provided, a 'hospital campus'. This contains more entries than a list of Trust codes, as one Trust may have services in more than one place. However it contains fewer entries than a full list of all hospital codes, as there may be many codes for services at the same physical location.

Analysts producing lists to specify the location of services for use in travel time work should do this by reviewing the campus list and using hospital codes, not Trust codes. Every site on their list should represent one physical campus with a known postcode.

Calculating the length of journeys

There are currently 3 different methodologies available that have been used in pilot projects to calculate the length of journeys between 2 postcodes. The advantages and disadvantages of each are described below. They range from simple calculations of straight line distances to more complex routing engines.

Work is underway to expand the number of methodologies available; the 3 main areas still in development are also described.

Available methodologies

Straight line distance (and scaled travel times)

The location of every postcode, in the eastings and northings coordinate system, can be found in the National Statistics Postcode Lookup. Once the location of the postcodes is known, the Euclidian straight line distance between them can be calculated.

It is possible to convert this distance into a crude travel time using a linear scale. Using a sample of known journeys, the formula:

$$time = 0.96 * distance \ + 9.3$$

(where distance is in km and time is in minutes), was derived using linear regression.

This methodology returns the same travel time for all journeys of the same distance, and so although it is indicative of travel time, more sophisticated methods are preferable. It would be unusual if it was the main methodology used for a project. It is useful for validating new methodologies, and providing an approximate travel time if a more sophisticated approach is not available.

ArcGIS travel times calculated by Norwich Medical School, University of East Anglia

Dr Peninah Murage of the University of East Anglia produced a table of travel times for journeys undertaken by private transport, which was shared with NCRAS. Travel times were calculated from all postcodes in England to all hospitals where there is a cancer MDT. The travel times were computed from the Spatial Analyst module using the 'Cost Distance' (impedance surface) command and the Meridian 2 road network in ESRI ArcGIS. Road speeds were taken from Jones et al. (2010)[3] and adjustments for walking speed were made for off road locations using the methodology set out in Bateman et al (2011)[4]. Full details are described in Sen et al. (2014)[5].

These travel times have been used in a range of studies, including the pilot study on survival. The limitations are that they were calculated for a set of cancer centres defined at one point in time, and as new centres open and services change the list will need updating. As they were part of an academic partnership, it is not possible for PHE to rerun these travel times independently, they depend on continued links with Dr Murage.

Graphhopper travel times

Graphhopper is a routing engine used by OpenStreetMap (www.graphhopper.com), which was used for the pilot study on prostate cancer treatment. Key to the decision to use Graphhopper was its availability as a Java library installable on a local machine, meaning that there is no need to send patient postcodes to a third party. It is open source, free and well used and tested. Using Graphhopper enables PHE to quickly rerun travel times when new start or end points are identified.

Graphhopper's main limitation is that the opensource free library contains no real world (GPS) time/speed data, which results in an underestimation of travel time of approximately 25%. This underestimation is largely consistent, so although the stated travel time is lower than experienced by patients, it appears robust for ranking travel times at a population level.

The routing engine could be run on public transport files in GTFS format[6], but has currently only been used to calculate distances by private transport.

Methodologies in development

Travel times calculated by the PHE GIS team

The PHE GIS team manages centralised spatial data and software on behalf of PHE and is working to develop precalculated road network and public transport travel time matrices on behalf of the organisation.

Esri’s **Network Analyst** extension is available as a function of Esri’s ArcGIS platform and allows analysis of road network travel distance and time: point to point, closest facility, catchment area analysis (isochrones).

TRACC is a software product developed by Basemap that enables public transport travel time calculations. It allows batch calculation of a number of different transport types: walking, cycling, driving and public transport. It is used by the Department for Transport for producing their journey time statistics, as well as other public bodies. Therefore it may help improve consistency across government-produced data. PHE currently has a TRACC licence and therefore could update calculations as service configurations change.

The software is proprietary and requires an annual licence. It also requires a server to run on, with associated costs and ICT support.

Both Network Analyst and TRACC are configurable to allow for a predefined methodology to be developed and implemented for routing analysis at PHE.

The central PHE GIS team currently use the OS Integrated Transport Network road network, but is reviewing options for using a more detailed road network.

The PHE GIS Team is currently working in collaboration with a travel times working group, ICT and Esri to understand the requirements for running large scale travel time analysis using PHE's High Performance Computing infrastructure. The aim is to produce a standard central set of travel times for both driving and public transport, using Network Analyst and TRACC respectively.

Google travel times

Google Maps is a high profile trusted routing engine widely used by many people to calculate travel times. An API exists to enable automatic calculation of travel times on the submission of a pair of postcodes, with a wide range of API clients supporting programming languages including Java and Python. Travel times are likely to be highly accurate as Google crowd sources traffic data, and there are options for modes of transport (including public transport), and adjusting times of day / traffic conditions.

The Google Maps travel times and supporting API are not open source. They support up to 2,500 requests a day for free, and charge for further requests. There have historically been contractual issues using Ordnance Survey data with the Google API. They also rely on the postcodes being sent to Google and the journey time being calculated remotely; it would not be possible to send them a list of postcodes of cancer patients because of the information governance risk of disclosure.

Making travel times available to all analysts

Not all analysts have access to routing engines like ArcGIS or Graphhopper. To increase access to travel times, a 'build once, use many times' approach was taken where the journey times were calculated for a specified list of pairs of postcodes (an embarkation postcode to a destination postcode). These calculations created a large pre-calculated table of travel times. This is specified as:

Embarkation postcode: the full set of UK postcodes (2,500,000 postcodes)

Destination postcode: to pre-calculate all travel times between all pairs of postcodes would result in a table of 6,250,000,000,000 rows. This is prohibitive to calculate and to store. The destination postcodes in the table are limited to postcodes of 271 locations identified as the main locations of cancer treatment for English cancer patients, which means only 680,000,000 rows need to be calculated. (These 271 locations were identified by working with clinical experts and with a data-driven list, with great thanks to Dr Murage whose original list of sites was the foundation for this work. New postcodes can be added to this list if new cancer treatment locations are identified.)

Distance: The straight line distance between the postcodes .

Regressed time: An approximate travel time linearly calculated from the straight line distance.

Norwich Medical School Times, UEA: The travel time between the postcodes as calculated by the partnership work with Dr Murage.

Graphhopper time: The travel time between the postcodes as calculated by the Graphhopper routing engine.

This format allows travel times from other sources and methods to be added to the table as they become available. It is then possible to rerun code with new travel time methodology by simply adjusting the columns selected for the calculated travel time. This structure also makes it easy to do comparative analysis of the different travel time methodologies.

Presentation of travel times

Average travel times of the cohort

An arithmetic mean of the travel times of all patients in a cohort can be calculated. This is easy to calculate, and can give quick comparisons between different groups of patients. For example, it may be useful for comparing the travel times of cancer patients in different geographical regions. If comparisons are being made between groups, providing confidence intervals around the average travel time is recommended.

This measure relies on the calculated travel times being accurate, and so may not be appropriate with travel times which are known to be underestimates. It also gives no measure of the distribution of travel times, and so does not distinguish between cohorts where everyone travels for 30 minutes and cohorts where half the patients travel for 10 minutes and half for 50.

Proportion of the cohort meeting a standard

For certain questions, there may be national standards or guidelines of interest, such as 'patients should not need to travel more than 45 minutes for this service'. Once travel times for the cohort of interest are calculated, the percentage who meet the standard can be calculated. Again, this measure relies on the calculated travel times being accurate, and so may not be appropriate with travel times which are known to be underestimates, and also gives no measure of the distribution of travel times. This measure is not recommended unless there is a compelling reason to publish it, and again should be caveated and presented with appropriate measures of uncertainty.

Distribution of travel times in a cohort

Instead of reducing the travel times to a single statistic as discussed above, a more informative way to present the data can be to plot the distribution of travel times for the cohort visually, usually as a line graph. It is also possible to produce a suite of statistics describing the distribution rather than one individual statistic, such as the mean, median, standard deviation, the inter quartile range, and other percentiles. A visual representation of the distribution of travel times can be a quick and intuitive way of communicating data. Again, if the travel times are known to be underestimates, caveats will be needed on interpreting this data.

Maps of travel times

Travel times to services of interest can be drawn on a specific type of map known as an isochrone map. This visualisation can be very powerful, and can clearly illustrate areas where travel times are longer than average. However, the majority of the population live in densely populated areas of the country, and presenting travel time information on a map can highlight the very long travel times of a small number of people in remote areas of low population density. This may miscommunicate a 'typical' travel time for the average person. A cartogram adjusting for population may resolve this, but at the expense of making the map more difficult to read.

Other statistics presented by travel time

Many travel time analyses aim to comment on whether a dependant statistic (for example, percentage of patients being treated with radiotherapy, one-year survival) varies with travel time. The usual way to present this is to divide up the cohort of all patients into several sub-cohorts by travel time, calculate the statistic for each of the sub-cohorts, and see if any significant difference is observed between the sub-cohorts.

The decision of how many sub-cohorts to divide the cohort into will depend on the size of the cohort. The larger the cohort, the more granular the division can be, which may aid in identifying trends, particularly non-linear ones. However, sub-cohorts should not be so small that the confidence intervals of the statistic of interest become large and statistical noise interferes with true trends.

Sub-cohorts can be defined by break points at fixed travel times (eg a sub-cohort of 'less than 10 minutes', another of '10 to 19 minutes' etc.). This makes interpreting any trend with travel time more natural. However, this may cause very small numbers in the sub-cohorts at the long right-hand tail of the data. An alternative is to divide the cohort into sub-cohorts by percentiles (eg quintiles). This keeps the numbers in all sub-cohorts equal, and avoids issues if the travel time engine is consistently underestimating. However, if there are effects that are only seen in the far right-hand tail of the data, use of large percentiles may miss these.

More sophisticated analysis may use travel time as a continuous variable in regression analysis, which avoids the need to identify sub-cohorts. However, it is still recommended that a basic calculation of the dependant variable by different travel time sub-cohorts is also produced, as observing the crude results visually by sub-cohort can greatly increase insight for building the correct model (if the trend is not linear, linear scaling such as linear regression may be inappropriate) and can aid with interpreting results.

Conclusion

The National Cancer Registration and Analysis Service as part of Public Health England has developed its understanding of the challenges of travel time calculations, particularly in the context of cancer care and National Health Service Trusts.

While a standard methodology is sought for routine publications, different research questions will require variations in the methodology used. Methodology should be tailored towards the question being asked. The current publications on travel times have a range of minor variations in the methodology used.

Identifying and defining cancer centres and specialist service providers remains a key challenge for travel time analysis. It is recommended that any project using a list of service providers transparently publishes this list, and that lists are reused where possible. It is acknowledged that minor changes in the time period covered or the question asked will require different lists of service providers.

Travel times to all major service providers have been pre-calculated, allowing data analysis by travel time without individual access to specialist software. Moving forwards, PHE continues to strive to improve the methodologies used to provide the most accurate and up-to-date travel times.

References

[1] Achieving world-class cancer outcomes: A strategy for England 2015-2020. Report of the Independent Cancer Taskforce
http://www.cancerresearchuk.org/sites/default/files/achieving_world-class_cancer_outcomes_-_a_strategy_for_england_2015-2020.pdf

[2] NHS Choices: https://www.nhs.uk/service-search

[3] Jones A, Wright J, Bateman I, Schaafsma M (2010). Estimating arrival numbers for informal recreation: a geographical approach and case study of British woodlands. Sustainability 2:684–701

[4] Bateman, I., Abson, D., Beaumont, N., Darnell, A., et al., “Chapter 22: Economic values from ecosystems” In UK National Ecosystem Assessment Technical Report. 1067–1152.

[5] Sen, A., Harwood, A.R., Bateman, I.J. et al. Environ Resource Econ (2014) 57: 233. https://doi.org/10.1007/s10640-013-9666-7

[6] GTFS format. The General Transit Feed Specification (GTFS) defines a common format for public transportation schedules and associated geographic information. GTFS "feeds" let public transit agencies publish their transit data and developers write applications that consume that data in an interoperable way:
https://developers.google.com/transit/gtfs/