HDRUK
Health Data Research UK

Exploring regional linked health data capability for research Phase 2: *Avoidable Attendances: Exploring variation in acute hospital admissions*

Contents

Executive Summary

- No existing national data feeds provide detailed and near-real time information on hospital admissions across the UK. **Currently available national data feeds are dated, do not include people still in hospital, and lack detailed coding which can help differentiate between different conditions or diagnoses.** Enabling a detailed, near-real time hospital admissions data feed of regional level data **would provide vital data for priority research and health and care planning.**
- This collaborative project builds on previous work (Phase 1) led by the Health Data Research UK (HDR UK) Regional Linked Health Data for Research Programme which aims to conduct '**driver' projects to explore data capability, data access and feasibility of enabling near real time hospital admissions linked data feeds at regional level.**
- **Phase 2** included 4 additional regions and implemented 2 driver use cases – this report summarises the driver use case led by the University of Sheffield which **explored variation in acute hospital admissions across the regions.**
- **Use Case Insights**: This driver use case identified that patients in the Emergency Care Dataset (ECDS) and Admitted Patient Care (APC) datasets were older in the Ambulatory Care Sensitive Conditions (ACSCs) groups on average than the non-ACSC groups. Deprivation was a key factor observed equally in ACSC and non-ACSC groups, and there was a high proportion of patients attending ED with ACSC. High variation existed between hospitals in terms of attendance for ACSC. Further research is needed to establish clearer criteria for potentially avoidable admissions and same day emergency care-eligible patients.
- **Data Capability and Access Insights** – Significant variance in data capability and access resulted in delays with clear opportunities for further harmonisation to promote more effective collaboration across multi regional data infrastructure.
- **Key Recommendations:**
  - **Recommend that regional linked data assets seek to obtain REC Research Database Approval**, enabling a more streamlined approach to multi regional collaborative projects
  - Guidance is needed to enable study teams to define multi regional projects sharing aggregated data only as **service evaluation or research.**
  - **Implementation of harmonised and standardised Data Sharing Agreements (DSAs) and Data Protection Impact Assessment (DPIAs) templates within and across regions** to enable streamlined multi regional collaborations.
  - **Ensure development of Reproducible Analytical Pipelines (RAPs) for multi-regional projects are developed at the outset** embedding robust principles of open science and code sharing
  - **Promote harmonisation of an agreed list of priority nationally and regionally collated specific data flows** across multi-regional TREs/SDEs (*e.g primary care, real time hospital admission discharge and transfer, secondary care laboratory and prescribing data)*
  - **Develop harmonised protocols across regions for Statistical Disclosure Control/ data minimisation**

Authors

Lead Site: University of Sheffield

- **University of Sheffield Study Team**: Suzanne Mason, Richard Jacques, Rebecca Simpson, Jen Lewis, Susan Croft, Madina Hasan, Simone Croft, Ross McMurray, Ric Campbell.

HDRUK Regional Linked Data Group

- **DataLoch (University of Edinburgh):** Nick Mills, Michael McDermott
- **iCARE /WISC (Imperial College Healthcare NHS Trust**): Eik Mayer, Ben Glampson, Catalina Carenzo
- **Barts Health NHS Trust:** Sophie Williams, Charles Gutteridge
- **University Hospitals Southampton NHS Trust:** Christopher Kipps, Rachel Chappell, Matt Stammers
- **University of Bristol:** Rachel Denholm, Amy Dillon
- **Lancashire Teaching Hospitals NHS Trust:** Jo Knight, Vishnu Chandrabalan, Quinta Ashcroft
- **PIONEER Health Data Research Hub:** Suzy Gallier, Rima Doal
- **Health Data Research UK:** Lara Edwards, Nada Karrar
- **NHS England:** Luke Readman

Introduction

The UK health data ecosystem has a significant gap that needs to be explored: the lack of detailed and near-real-time information on all hospital admissions across the country. Currently, the national data feeds available are outdated, with up to a six-week time delay post-discharge, and do not provide diagnostic granularity or include people still in hospital. This gap can be addressed by implementing a detailed, near-real-time hospital admissions data feed that would provide vital data for research and NHS clinical audit programmes contributing to the development and evaluation of the safety and effectiveness of clinical pathways, precision medicine and the application of computational methods for healthcare use.

This is not only critical for current high-priority pandemic research, but also for preparing for future healthcare system threats. By building a real-time sentinel, data-driven surveillance system for National Healthcare Systems, the quality and timeliness of data available for critical research and analysis would improve, better equipping healthcare providers and researchers to respond to healthcare crises in a timely and effective manner.

Phase 1

In 2021, under the auspices of the Data and Connectivity National Core Study (D&C NCS), HDR UK, Luke Readman (Director of Digital Transformation, NHS England) convened a collaborative group across five health and care regions across England and Scotland.

Using COVID-19 vaccine related thrombotic adverse events as a use case (D&C NCS Use Case One), the group explored the feasibility of enabling regional, rapid near real-time acute admissions data flows and of scaling up across the UK. The full report[1] of this use case can be accessed here.

This work demonstrated key differences across the five ICS in terms of established data availability and processes and presented the following recommendations for consideration:

1. **Identify barriers for Integrated Care Systems (ICS) to use SNOMED CT coding at point of care.** Mandate standardised point of care coding using SNOMED CT, starting with digitally mature ICS and expanding out nationally, and set local incentives for reporting.
2. **Ensure each ICS receives timely data from the national datasets to which they contribute - data availability has caused delays in progress**, (e.g., lack of national vaccination data available for PIONEER). We recommend that all ICSs are sent regional data slices for their population from national datasets to carry out high priority research.
3. **Harmonised governance processes across regional and national data custodians** – e.g., implement a core set of 'approvals for tasks' expected for each data custodian and then a national system to rapidly approve with targets from application to approval.
4. **Fund robust public involvement and engagement** to ensure transparency, identify concerns and build public trust, particularly around access and privacy.
5. **Develop and leverage regional expertise embedded within each ICS** - building on existing relationships between population health management groups, analysts and researchers.

[1] https://www.hdruk.ac.uk/wp-content/uploads/2023/06/Regional-Linked-Health-Data-Report_-Phase-1-VITT-use-case_FINAL.pdf

Phase 2

In April 2022, the regional model was scaled up to include four additional regions:

- University Hospital Southampton NHS Foundation Trust
- University of Sheffield
- Lancashire Teaching Hospitals NHS Foundation Trust
- University of Bristol

The group has continued to explore the data capability, variances in IG and data access across different regions through the implementation of two further use cases:

- **Exploring variation in acute hospital admissions**. Developing criteria to define an avoidable admission: Principal Investigator - Professor Suzanne Mason, University of Sheffield (Use Case Two). *See Appendix 1 for Study Protocol.*
- **Evaluation of the effectiveness and safety of early rule out pathways for acute myocardial infarction**: Principal Investigator - Professor Nicholas Mills, University of Edinburgh (Use Case Three)

This report describes the finding of use case two - *Exploring variation in acute hospital admissions,* with key learnings, insights and recommendations from this phase included in the discussion section.

Characteristics of Phase 2 regions

Analyses were performed across eight regions: Barts Health NHS Trust (Northeast London), iCARE/WSIC (Imperial College Healthcare NHS Trust, Northwest London), PIONEER HDR Data Hub (Birmingham and West Midlands), DataLoch (Southeast Scotland), University of Sheffield (Yorkshire and Humber), Lancashire Teaching Hospitals NHS Foundation Trust (Lancashire and South Cumbria), University Hospital Southampton NHS Foundation Trust (Wessex) and University of Bristol. The table below describes the characteristics of each region – including the Integrated Care Board (ICB) or Health Board (Scotland) that each regional collaborator is part of. Data from NHS Trusts within respective ICS/Health Boards were included in the analysis. The majority of regions accessed and analysed data within regional Trusted Research Environments (TREs) /Secure Data Environments (SDEs)[2], with some (e.g Barts) receiving anonymised data extract via local IG processes and approvals.

[2] https://www.hdruk.ac.uk/access-to-health-data/trusted-research-environments/

Table 1 and 2: Regional characteristics

	PIONEER	iCARE	Dataloch	Barts
ICB or Health Board	NHS Birmingham and Solihull ICB	NHS Northwest London ICB	NHS Lothian Health Board	NHS Northeast London ICB
Population size (millions)	1.3m	2.1m	0.85m	2.0m
Research or service evaluation?	Research	Research	Research	Research
Is research ethics database approval in place?	Yes	Yes	Yes	No: Joint Research Office approval required
IG process summary	1. Data Request Form submitted 2. Data Trust Committee approval 3. Data Controller final approval 4. Data licence is drafted and access granted when signed	1. Data access form submitted 2. Applications are triaged by clinicians, informatics, Data Protection, and PPIE leads 3. Data Access Committee presentation and approval (Community partners/Caldicott Guardian/Data Protection Office/Healthcare Professionals/Joint Research Office/Academics 4. Training, induction & on-boarding to the SDE	1. For analysis of row level data, research protocol submitted 2. Ethics panel, patient and public ref group and Caldicott Guardian review and approval 3. Rapid access to external researchers via eDRIS	1. Request for retrospective anonymised data accessed by trust submitted 2. Data Access Committee review 3. If threshold is met, DPIA will be completed and DPO have final sign off 4. If threshold is not met, JRMO will complete all further checks 5. IG team also reviews all DPAs and DSAs 6. DPIA was not required in this project
Primary care data – availability and linkage capability	No	Yes	Yes	Yes
Data access via TRE/SDE?	Yes	Yes	Yes	No

	Sheffield	Southampton	Lancashire	Bristol
ICB or Health Board	NHS South Yorkshire ICB	NHS Hampshire and Isle of Wight ICB	NHS Lancashire and South Cumbria ICB	NHS Bristol, North Somerset & South Gloucestershire ICB
Population size (millions)	1.4m	1.4m	1.8m	1.0m
Research or service evaluation?	Service evaluation	Service evaluation	Service evaluation	Service evaluation
Is research ethics database approval in place?	No: University of Sheffield ethics committee approval	Not yet - in progress	Not yet - in progress	No

IG process summary	Caldicott Guardian approval required DPIA submitted for each hospital site (x 4) DSA submitted for each site (3 sites used UoS template) IG access is approved by multidisciplinary panel	Caldicott Guardian approval required DPIA submitted DPIA reviewed by IG team and CG for final approval	Service evaluation registration request submitted Application reviewed by Head of R&I for ethical issues Input from IG team may be requested but not in this case, as it was anonymized data not going outside the team	Unlinked Trust Data (for each Trust): Caldicott Guardian Approval, Collaboration Agreement, Data Transfer Agreement, DPIA Linked Data (with bespoke 2xTrust extracts): Trust Caldicott Guardian approval, Trust DPIAs, NHS Standard Contract schedule 6 specifications and approval, Statements of work (including specifications) for Trusted Third Party, ICB viability check and approval, GP opt-out, Data processing agreement with embedded specification.
Primary care data – availability and linkage capability	No	No	No	Yes
Data access via TRE/SDE?	No	Yes	Yes	No

Each region within the group applied to access their local data via existing Information and Research Governance processes. Aggregated data subject to site statistical disclosure control was sent to the two lead sites (Sheffield and Dataloch) for collation and summary reporting.

The project has provided detail on the different data access and governance processes in place across different regions for research and service improvement. For example, some regions have existing CAG and ethical approval in place as a HRA research database for agreed scope of research (e.g. PIONEER, Imperial (iCARE)) and others needed to apply for local ethics and IG approvals on a per project basis.

In addition, some sites classified the local analysis as service improvement, others as research. Overall, this work has demonstrated the challenges for research projects wishing to access regional level linked health data in more than one region, as there is much variation in data access and governance processes.

Analytical Report

Exploring variation in acute hospital admissions. Developing criteria to define an avoidable admission:

Background

- There has been a large increase in waiting times for emergency care across the country. It is not clear why this is happening, and in order to understand how we can solve the problem, we need to understand this demand and where it is coming from.

- The National CORE 20 Plus 5 priorities[3] is about reducing healthcare inequalities and place-based approaches are essential to understand – so getting a greater regional / subregional view of how we are handling these priorities is important. Under the priorities there is a drive from NHSE to see a shift in both inequalities in unplanned hospital admission for chronic Ambulatory Care Sensitive Conditions (106a) and Urgent Care Sensitive Conditions (106b).
- The latest urgent and emergency care strategy focuses on reducing unwarranted variation in systems as well as expanding services in the community to manage winter pressures. The ability to analyse the impact of service changes introduced is vital if the NHS is to learn and optimise delivery.
- National bed occupancy rates are consistently higher than the recommended 85% and strategies are needed to reduce them. A Monitor report done in 2015 identified bedded capacity as the key bottleneck for deteriorating A&E performance.
- There is a defined list of ambulatory care sensitive conditions with ICD10 codes which means identifying avoidable admissions is possible using this list as an initial approach to analysis.

Aim of Study

- To understand variation across the country in all acute hospital admissions and ED attendances
- To explore methods for identifying an avoidable acute admission focussing on ambulatory care sensitive conditions

Ambulatory care sensitive conditions (ACSCs) are conditions where effective community care and case management can help prevent the need for hospital admission.

Objectives

- Take a multi-regional approach to analysing routine real world health data to describe patterns of acute admission and regional variation in admissions over a defined period of time.
- Analyse variation between sites in acute admission and ED attendance in ACSC by factors including deprivation, patient demographics, time of day, day of week, waiting times.
- Use hospital admission data to describe outcomes following acute admission using ACSC codes from the Emergency Department and also acute admissions data

Data Sources Analysed

- **Emergency Department Data**

  All adult (18 years and above on day of attendance) with an unplanned first emergency care attendance for a new clinical condition (or deterioration of a chronic condition) at Type 1 Emergency Departments (Emergency Departments that are consultant led 24-hour service with full resuscitation facilities and designated accommodation for the reception of

[3] https://www.england.nhs.uk/about/equality/equality-hub/national-healthcare-inequalities-improvement-programme/core20plus5/

emergency care patients) within the study time period (1st November 2021 to 31st October 2022 inclusive).

To ensure consistent analysis data sets were used across regions/sites we attempted to align the data specification for the Emergency Department data to the nationally (England) mandated Commissioning Data Set 6.2 Type 011: Emergency Care[4] and its format or coding..

The data specification requested four types of variables. Filtering variables for identifying attendances to be included in the study, and patient characteristics/demographics, attendance characteristics and attendance outcomes used for producing aggregated analyses.

Filtering Variables

- ED Department Type: used to select Type 1 Emergency Departments
- Attendance Category: used to identify unplanned first emergency care attendances
- ED Arrival Date/Time: used to identify attendances within the study time period

Patient Characteristics/Demographics

- Age at time of activity (also used as a filtering variable)
- Gender
- Ethnicity (groups defined in the 2001 census)
- Townsend Score Quintile: a measure of deprivation derived from patient postcode that cover both England and Scotland[5]
- Accommodation Status (used to identify patients who live in residential or nursing homes)
- Comorbidities

Attendance Characteristics

- Provider Code: the NHS Digital ODS code defining the organisation providing treatment
- ED Site Code: the site code for each Type 1 Emergency Department at the Trust
- Arrival Mode (Walk-In, Ambulance, Other)
- Source of Attendance (e.g. Emergency Services, Hospital, Primary Care)
- Chief Complaint: the chief complaint as assessed by the care professional first assessing the patient
- Acuity: the acuity of the patient's condition at the time of initial assessment
- Diagnosis: used to identify Ambulatory Care Sensitive Conditions
- Diagnosis Qualifier: the level of certainty of a patient diagnosis
- Time in ED: the time in minutes that a patient spent in ED
- Investigations: clinical investigations performed while a patient is under the care of the Emergency Department
- Treatments: treatments performed while a patient is under the care of the Emergency Department

Attendance Outcomes

[4] https://www.datadictionary.nhs.uk/data_sets/cds_v6-2/cds_v6-2-3_type_011_-_emergency_care_cds.html#dataset_cds_v6-2-3_type_011_-_emergency_care_cds

[5] https://statistics.ukdataservice.ac.uk/dataset/2011-uk-townsend-deprivation-score)

- Discharge Destination (used to identify patients who are discharged, admitted, died, etc)
- Referred to Service (Inpatient service to which the patient was referred for admission or opinion by treating clinician)

- **Admitted Patient Care Data**

  Patient records for all completed acute emergency admissions for adults (18 years and above on day of admission) that had an admission date within the study time period (1st November 2021 to 31st October 2022 inclusive).

  To ensure consistent analysis data sets were used across regions/sites we attempted to align the data specification for Admitted Patient Care data to the nationally (England) mandated Commissioning Data Set 6.2 Type 130 Admitted Patient Care - Finished General Episode[6] and its format or coding.

  The data specification requested four types of variables. Filtering variables for identifying admissions to be included in the study, and patient characteristics/demographics, admission characteristics and admission outcomes used for producing aggregated analyses.

  Filtering Variables

  - Admission Method: used to identity acute emergency admissions
  - Admission date: the date the patient was admitted to hospital at the start of a hospital spell, used to identify attendances within the study time period
  - Episode Number: the order of the episode within the current hospital provider spell, used to obtain the diagnosis and procedure data from only the first episode
  - Spell Number: a unique identifier for each hospital provider spell (used to consolidate episodes into spells)

  Patient Characteristics/Demographics

  - Age on admission (also used as a filtering variable)
  - Gender
  - Ethnicity
  - Townsend Score Quintile

  Admission Characteristics

  - Provider Code: the NHS Digital ODS code defining the organisation providing treatment
  - Hospital Site Code: the site code for hospital of admission
  - Source of Admission: used to identify possible care home admissions
  - Primary Diagnosis: used to identify ambulatory care sensitive conditions (the primary diagnosis from the first episode in a hospital provider spell was requested)
  - Secondary Diagnoses: used to identify comorbidities (secondary diagnoses from the first episode in a hospital provider spell were requested)

  Admission Outcomes

[6] https://www.datadictionary.nhs.uk/data_sets/cds_v6-2/cds_v6-2_type_130_-_admitted_patient_care_-_finished_general_episode_cds.html

- Length of Stay: the length of stay of the patient admission, calculated as the time between a patient's spell start data and spell end date
- Discharge Destination: indicates where the patient was due to go on leaving hospital (can be used to identify discharge to care homes)
- Discharge Method: indicates the circumstances under which a patient left hospital (can be used to identify deaths)
- Procedures: operative procedure codes recorded for the episode

Methods

The approach to data collection, validation, processing/feature engineering and analysis varied by site.

1. **Data Collection:**
   - Prior to data extraction, the lead site circulated a data specification document to all other sites so that they could feedback on data availability.
   - Raw data was extracted either (a) directly from databases by analysts working at sites where existing data pipelines were in place, or (b) by hospital trust business intelligence analysts prior to secure transfer to sites for processing/feature engineering.
   - Data included patient records for first time emergency care attendances and acute admissions with the specifications mentioned above.
2. **Data Harmonisation: Validation and Feature Engineering**

   For each site to produce a standard, consistent and processed data set that can be analysed to produce reliable outputs, a collaborative data harmonisation pipeline was created by analysts. The lead site created a data processing document that described all variables in the data specification and any recoding required (for example, how age should be recoded into categories).

   - **Data validator and feature engineering pipeline:**
     - A data validation and feature engineering pipeline was collaboratively developed and is accessible here: https://github.com/LTHTR-DST/hdruk_avoidable_admissions.
     - Associated documentation was written to assist utilisation of the Python package: https://lthtr-dst.github.io/hdruk_avoidable_admissions/ and general collaboration https://mattstammers.github.io/hdruk_avoidable_admissions_collaboration_docs/
     - A method of mapping SNOMED-CT terminology was integrated into the Python module
     - Not all sites were able to use the data validator and instead developed their own validation and data processing procedures.

3. **Statistical Analysis Plan:**

   A Statistical Analysis Plan and Analysis Tables document were developed by the lead site prior to receiving any data. These documents detailed the methods for producing aggregated data.

The analysis plan was based on simple data summaries in order not to overburden sites with complex analysis or excessive numbers of aggregated data tables.

A brief summary of the analysis is as follows:

- For both the ECDS (ED) dataset and the APC dataset, summary statistics of the patient demographics, attendance characteristics and attendance outcomes were calculated.
- For numerical variables, the minimum (min), maximum (max), mean, standard deviation (SD), median, lower quartile (Q1) and upper quartile (Q3) were presented with the number of observations used in the calculations.
- For categorical variables, the number and percentage of patients in each of the categories and the total number of observations were calculated.
- These summary statistics were broken down by patients who attended with an ACSC and those who attend with other conditions.
- All analyses were split by hospital site level.

4. Data Extraction

- The anonymised, aggregate data tables created by each site within local environments were released and sent to the lead site.
- The lead site defined a single rule for the suppression of small numbers, taking into account the details of the Statistical Disclosure Control (SDC) agreed by each site, for those sites who required to apply when sending the aggregated data because of local restrictions. For those who did not need to suppress small numbers according to their specific ethics or data agreements, the actual numbers were sent to the lead site, who then performed the SCD.
- The agreed rules for suppression of small numbers were as follows:
    - Only suppress at the final stage of analysis (i.e. after the aggregation of categories described in the data processing document)
    - For cells with N = 0, report 0
    - For cells with 0 < N < 10, report as -1 (chosen to be distinct from any possible numbers in the data)
    - Exclude -1 values from the calculation of percentages and report as NA
    - After suppression of small numbers round all other numbers to the nearest 5

5. Aggregated Data Analysis

- Aggregated data from individual sites was stored, processed and analysed in the lead site's Secure Data Environment
- Aggregated tables were processed to ensure consistent formatting prior to creation of final analysis data sets
- Aggregated data was summarised at a hospital and overall level to meet the objectives set out above

Results

As of 30 April 2023 the lead site had received data from 7 sites covering 12 hospital trusts, 20 type 1 emergency departments with a total of 1.4 million first time attendances (median: 71,175, range: 23,880 to 114,190) and 23 general (non-specialist) hospitals with a total of 575,000 acute admissions (median: 23,500, range: 3,960 to 51,250).

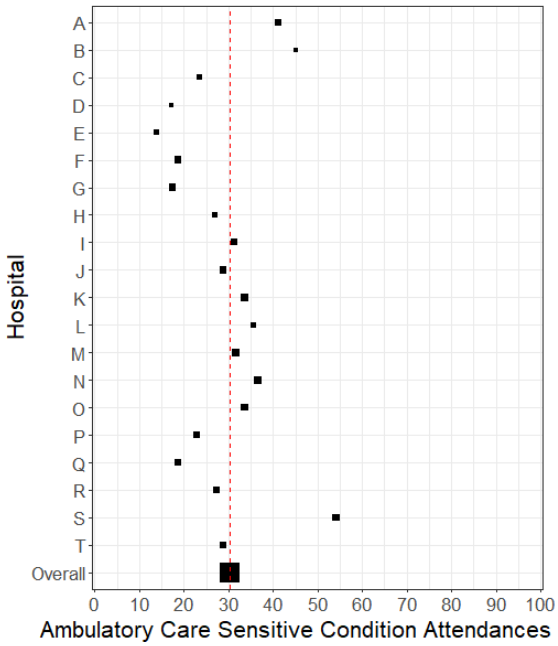
Emergency Care Dataset (ECDS)

Ambulatory Care Sensitive Conditions (ACSC) presenting at the Emergency Department

Figure AA1 shows the percentage of ED attendances that were for ACSC. Overall, 30.2% (423,235/1,399,940) of first time attendances were for ACSC but there is significant variation between different hospitals with the percentage ranging from 13.8% (7,350/53,095) to 54.2% (48,795/90,040). The frequency of attendances by type of ACSC is shown in Figure AA2.

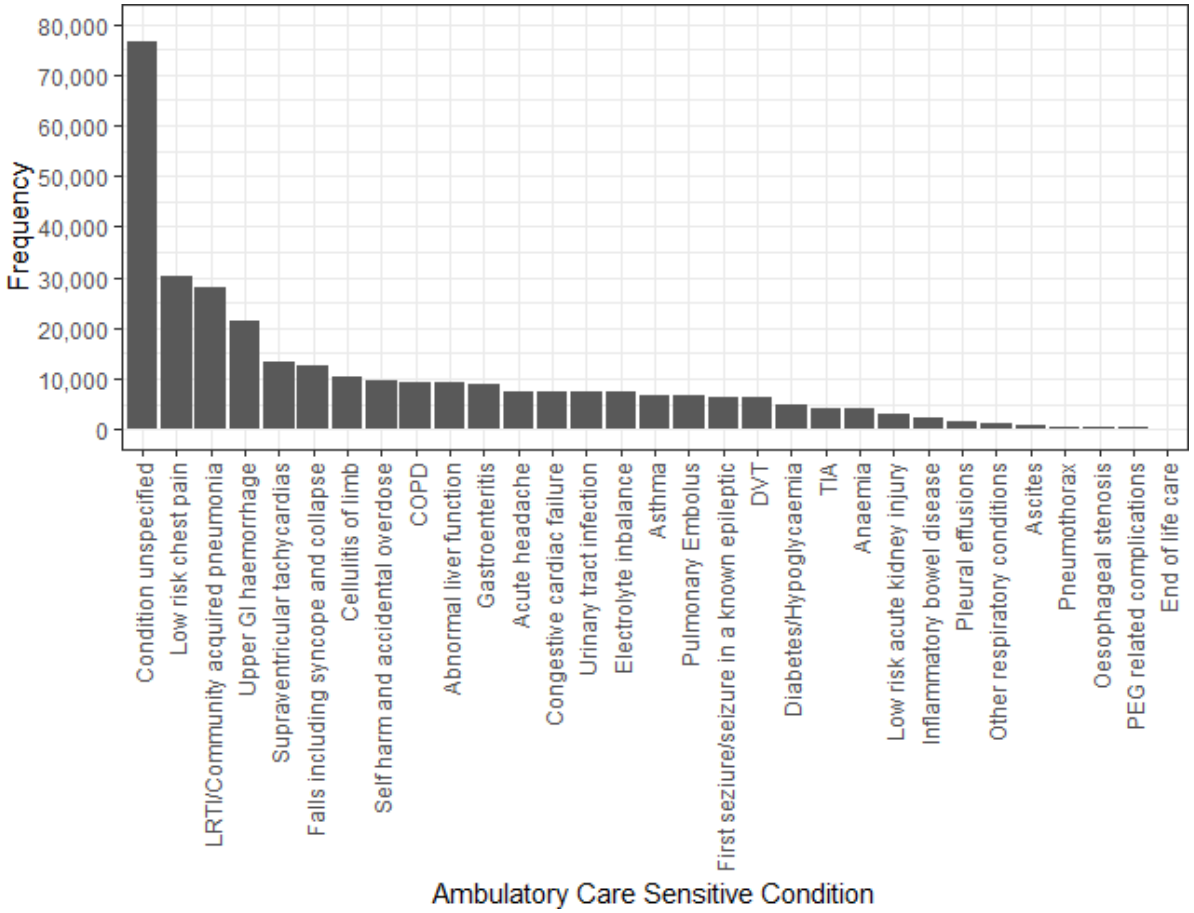
The most common conditions (based on data from 15 out of 20 hospitals) are 'condition unspecified' (76,695 attendances, 25.9% of ACSC), low risk chest pain (30,290 attendances, 9.8% of ACSC), lower respiratory tract infection or community acquired pneumonia (28,080 attendances, 9.1% of ACSC) and upper gastrointestinal haemorrhage (21,290 attendances, 6.9%). These four conditions account for 50% of all ACSC attendances.

Figure AA1: Percentage of ED Attendances for ACSC



	ACSC Attendances	All Attendances	Percent ACSC
A	34510	83870	41.1
B	15185	33690	45.1
C	11915	50735	23.5
D	5655	32880	17.2
E	7350	53095	13.8
F	13710	73125	18.7
G	12340	70750	17.4
H	11835	44025	26.9
I	21485	68600	31.3
J	29270	101535	28.8
K	38505	114190	33.7
L	15470	43365	35.7
M	27810	87905	31.6
N	36650	99955	36.7
O	26735	79580	33.6
P	16970	74080	22.9
Q	13250	71175	18.6
R	15540	56965	27.3
S	48795	90040	54.2
T	20255	70380	28.8
Overall	423235	1399940	30.2

Figure AA2: Frequency of ED attendances by ACSC



Patient Demographics

The age profile of patients attending with ACSCs differs to those attending with other conditions (Figure AA3). The highest proportion of attendances for ACSCs are in the 85 and over age group (9.2%, 38,115/423,235) and the majority of attendances for ACSCs are 50 and over (57%, 241,235/423,235). In comparison, the highest proportion of attendances for non-ACSC is in the 25-29 age group (9.8%, 95,335/976,665) with the majority of these attendances aged under 50 (54.6%, 533,630/976,665). Figure AA4 shows the age distributions of ACSCs and other conditions by hospital. Although there is some variation in the median age by hospital the differences between groups of conditions follows a similar pattern with attendances for ACSCs being on average older than those for other conditions.

Figure AA3: Percentage of attendances for ACSCs and Other Conditions by age group

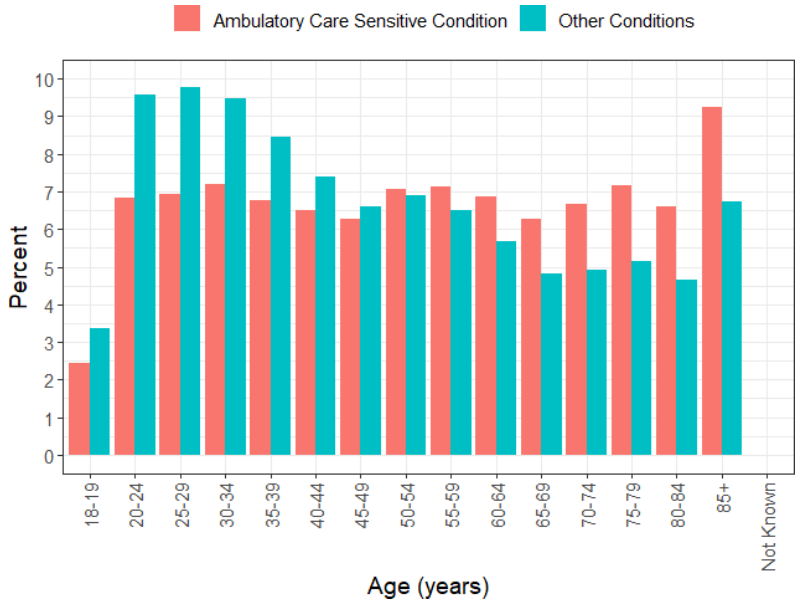
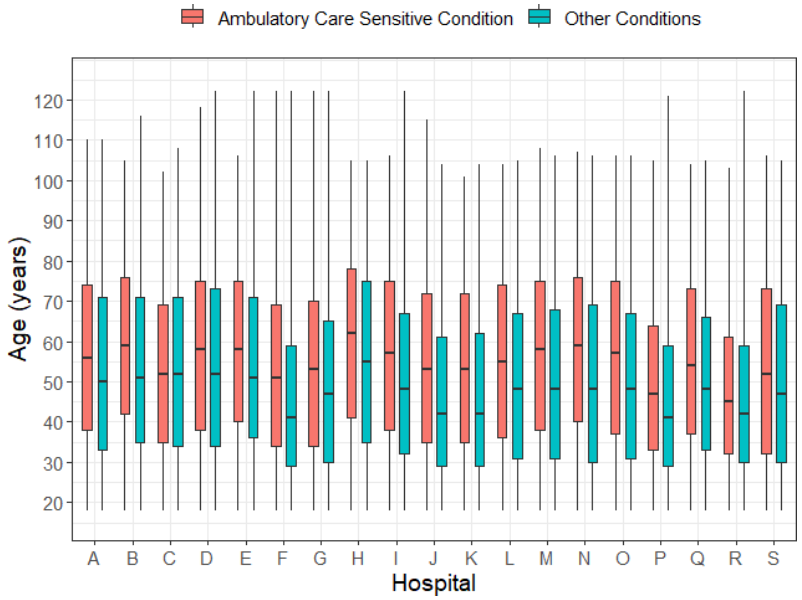
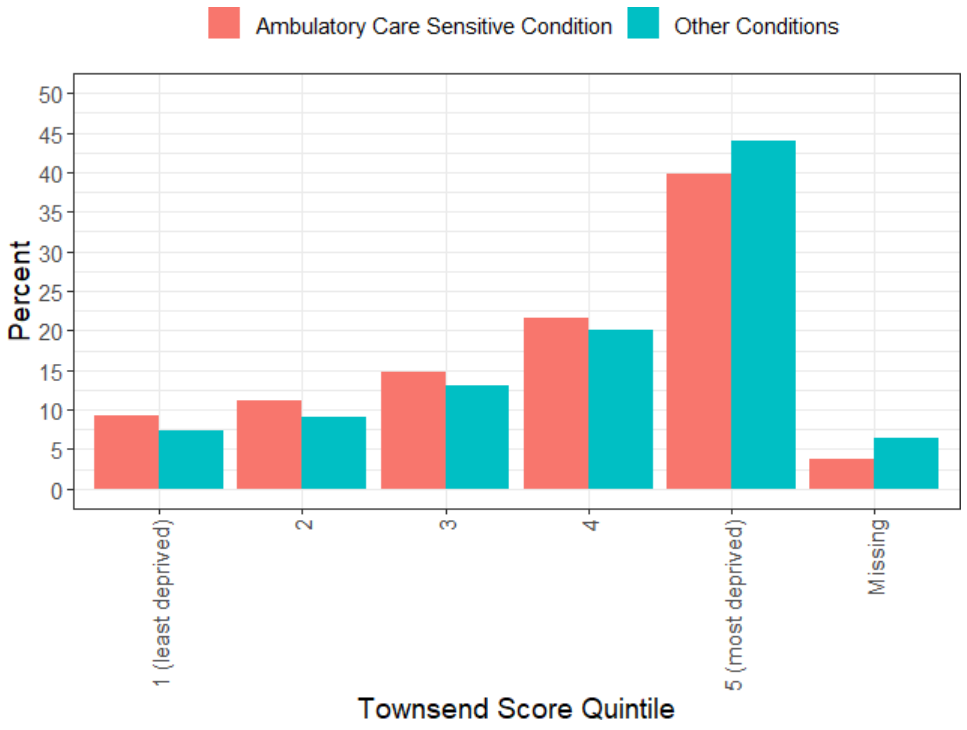


Figure AA4: Boxplots of age by type of condition (ACSC versus other) and hospital



Deprivation, measured using the Townsend deprivation index at the lower layer super output area level, shows similar patterns for both ACSCs and other conditions (Figure AA5). For groups both of conditions the largest proportion of attendances are in the most deprived quintile (39.7% (167,285/421,140) of ACSCs verses 44.0% (427,065/970,455) of non-ACSC) and the proportion of attendances increases monotonically from the quintile of least deprived to the quintile of most deprived.

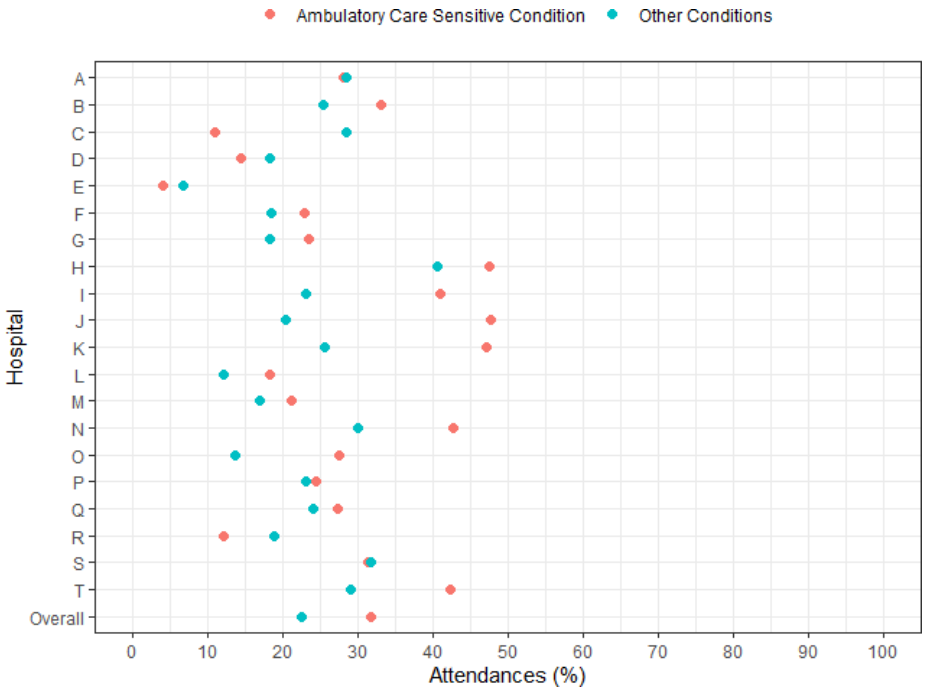
Figure AA5: Percentage of attendances for ACSCs and Other Conditions by Townsend Score Quintile of Deprivation



Discharge Destination

The discharge destination of patients attending with ACSC differs to those with other conditions. Overall, the largest differences were in attendances that resulted in admission to hospital or discharge with 31% of attendances for an ACSC being admitted (60% discharged) compared to 23% (66% discharged) of non-ACSC attendances. Smaller differences were seen in other discharge destinations: died (0.1% of ACSC vs 0.2% of non-ACSC), ambulatory/short stay (2.9% of ACSC vs 4.3% of non-ACSC), and transferred (0.5% of ACSC vs 3% of non-ACSC). There is significant variation in the percentage of attendances for ACSCs that result in admission between hospitals from 4% to 48% (Figure AA6). It is also worth noting that for the majority of hospitals (14/20) the proportion of attendances for an ACSC resulting in admission is larger than the proportion for non-ACSC.

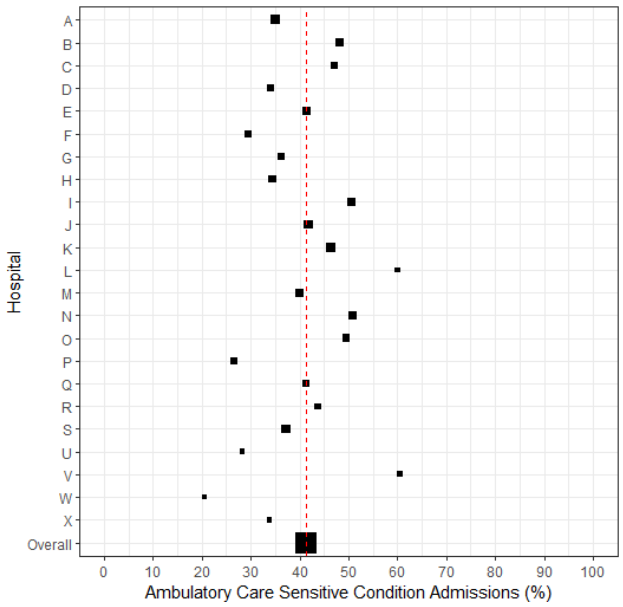
Figure AA6: Percentage of attendances for ACSC and other conditions resulting in admission by hospital and overall


Ambulatory Care Sensitive Condition
Other Conditions
Hospital
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
Overall
0
10
20
30
40
50
60
70
80
90
100
Attendances (%)

Hospitals Admission Dataset (Admitted Patient Care Data)

The overall percentage of patients being admitted with ACSC was 41.2% (236,615/574,650), again with considerable variation between hospital sites which have a range from 20.5% to 60.6%. (See Figure APC1)

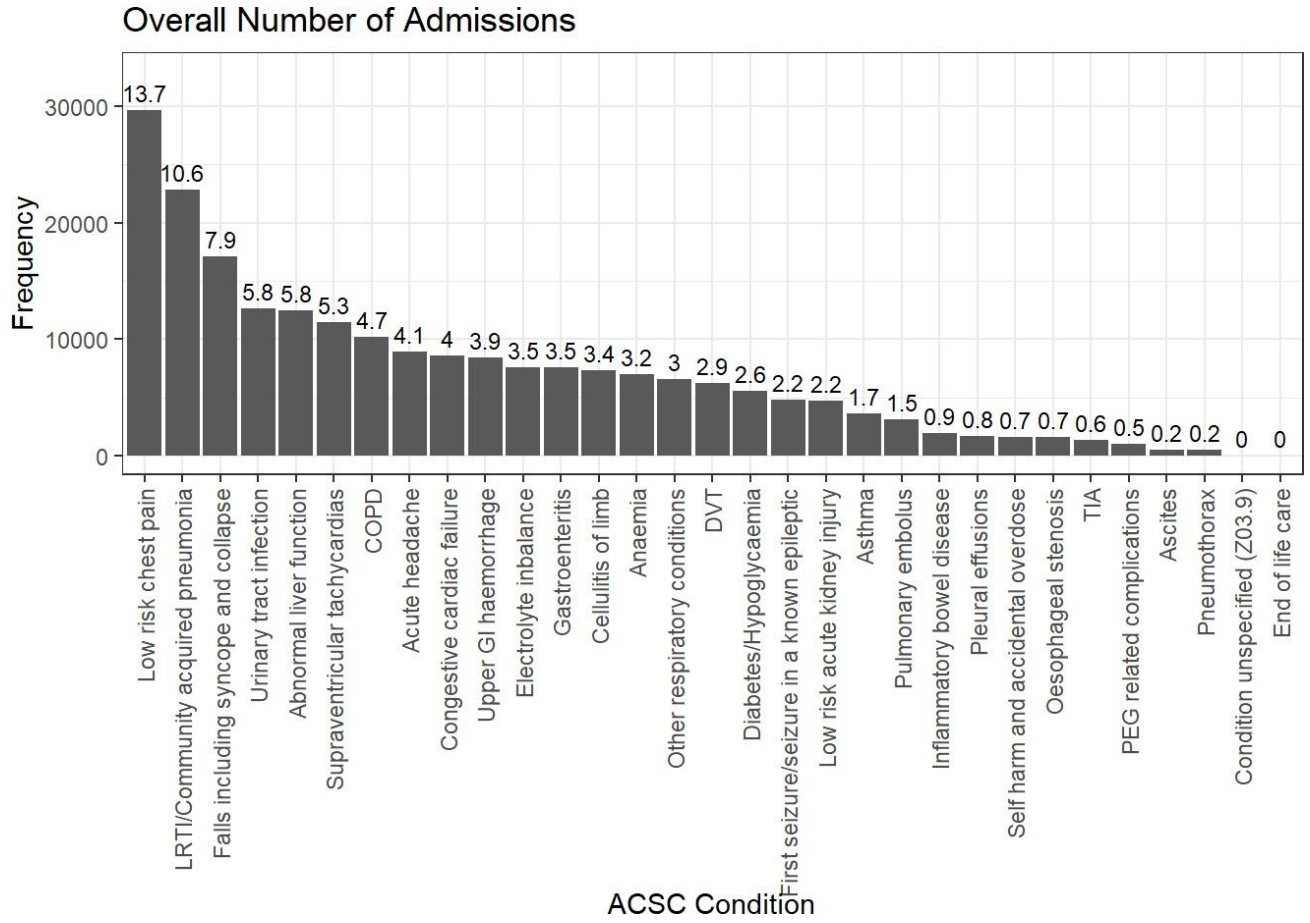
Figure APC1: Percentage of hospital admissions for ACSC



ACSC Attendances	All Attendances	Percent ACSC
17940	51250	35
12955	26865	48.2
9210	19570	47.1
7380	21650	34.1
12670	30610	41.4
5270	17935	29.4
5620	15480	36.3
10860	31540	34.4
17550	34635	50.7
21195	50880	41.7
23640	50990	46.4
3660	6100	60
12565	31530	39.9
15490	30425	50.9
14600	29515	49.5
6250	23500	26.6
7365	17780	41.4
4635	10605	43.7
18910	51015	37.1
1735	6125	28.3
4565	7535	60.6
810	3960	20.5
1740	5155	33.8
236615	574650	41.2

Figure APC2 presents the breakdown of ACSC conditions for all ACSC admissions The most common conditions were low risk chest pain (N=29,690, 13.7%), lower respiratory tract infection or community acquired pneumonia (N=22,865, 10.6%), Falls including syncope and collapse (N=17,105, 7.9%), Urinary tract infection (N=12,615, 5.8%) and Abnormal liver function (N=12,455, 5.8%). These top 5 conditions account for just under 45% of all ACSC admissions to hospital.

Figure APC2-overall admissions to hospital by diagnosis



Patient Demographics

Older people are more likely to be admitted with an ACSC condition with those in the eldest category (85+) having the highest proportion (N=34,925, 14.8%) of ACSC admissions. However, those in the eldest category also had the highest proportion (N=38,035, 11.3%) of non-ACSC admissions. (Figure APC3)

Figure APC3: Admissions by age group

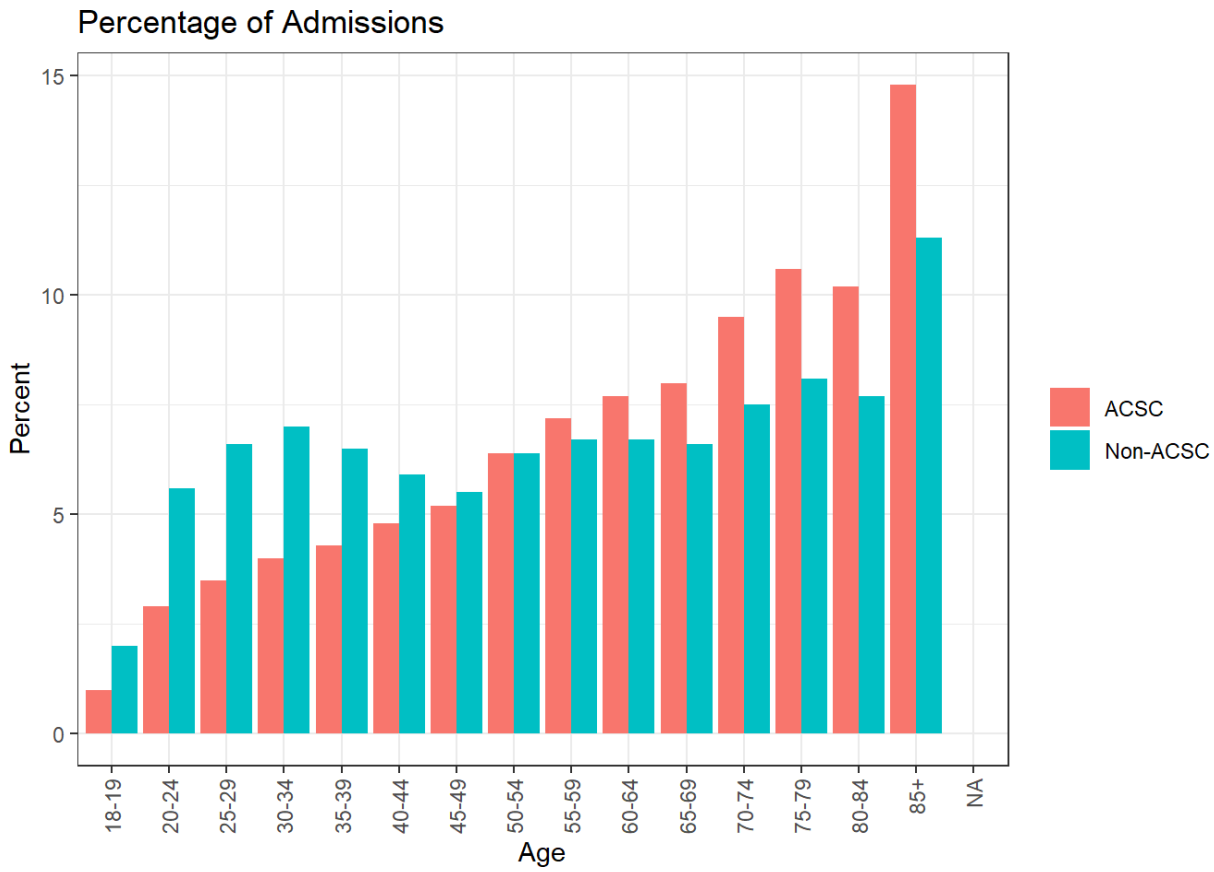


Figure APC4 below presents the age distributions for ACSC and non-ACSC split by hospital site. There is some variation in the age distributions, however for the majority of the sites those who are admitted for an ACSC tend to be older.

Figure APC4: Variation in age at hospital admission by site

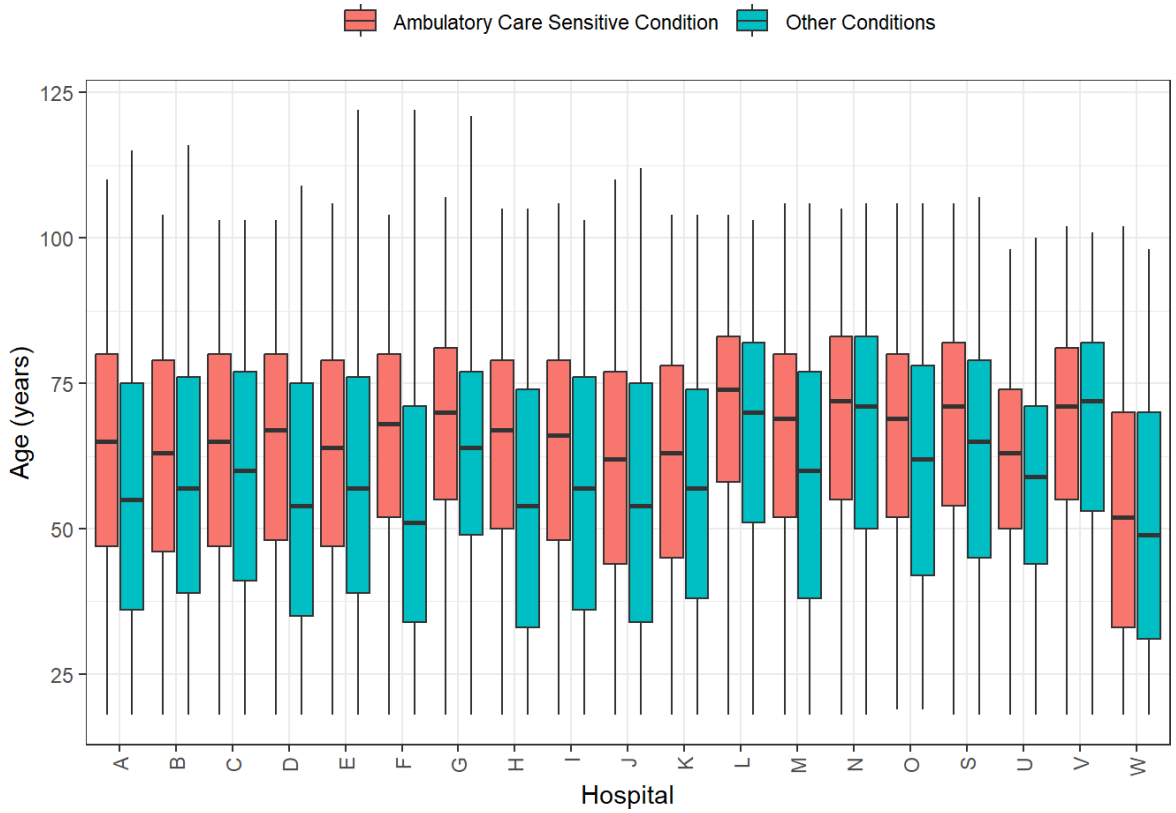
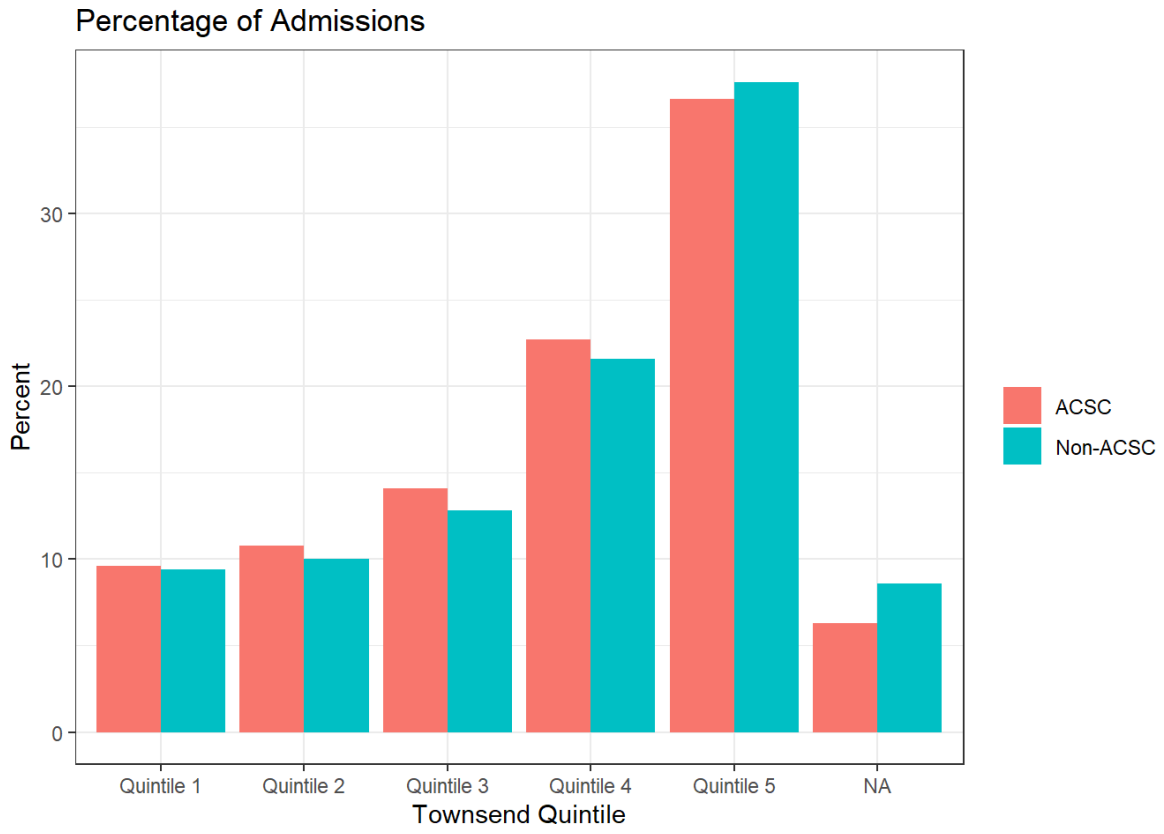


Figure APC5 shows the largest proportion of admissions (for both ACSC (N=63,495, 36.6%) and non-ACSC (N=98,795, 37.6%)) were in the most deprived quintile (5). The plot also shows that the proportion of attendances increases monotonically from the quintile of least deprived (1) to the quintile of most deprived (5).

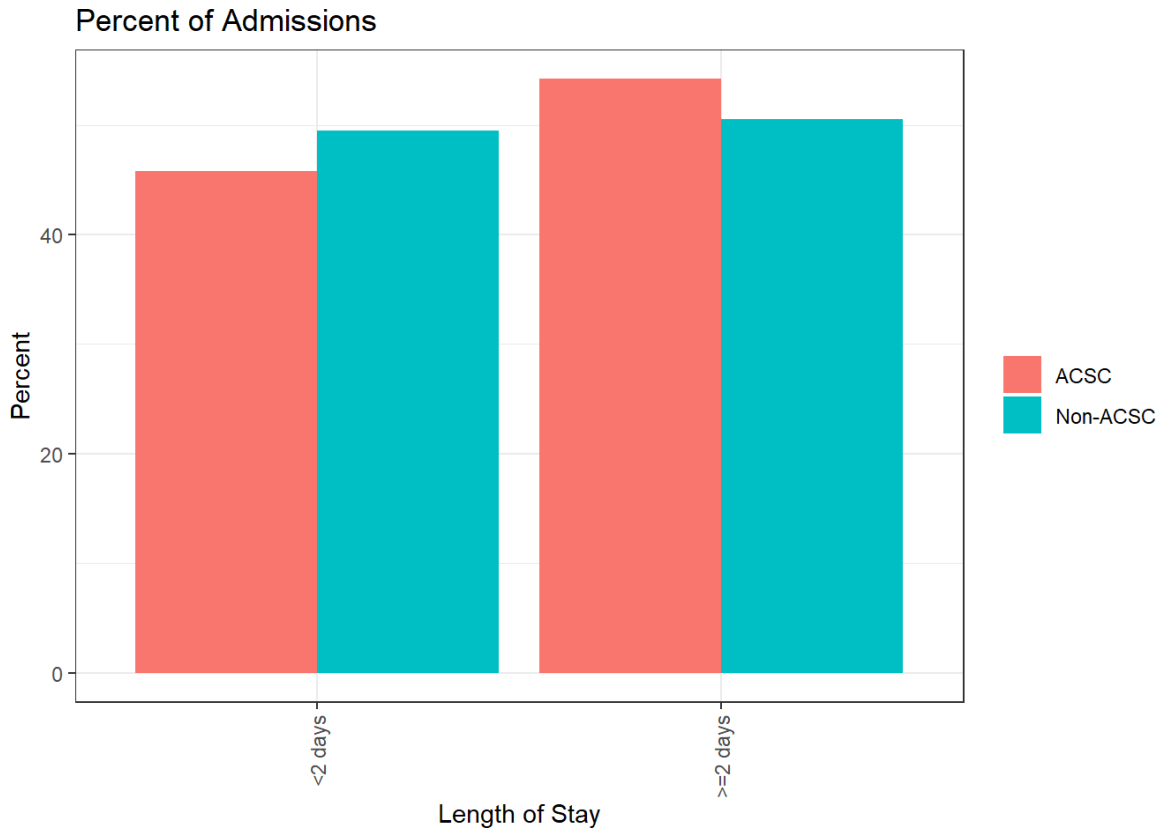
Figure APC5: Hospital admissions by deprivation



Attendance outcomes

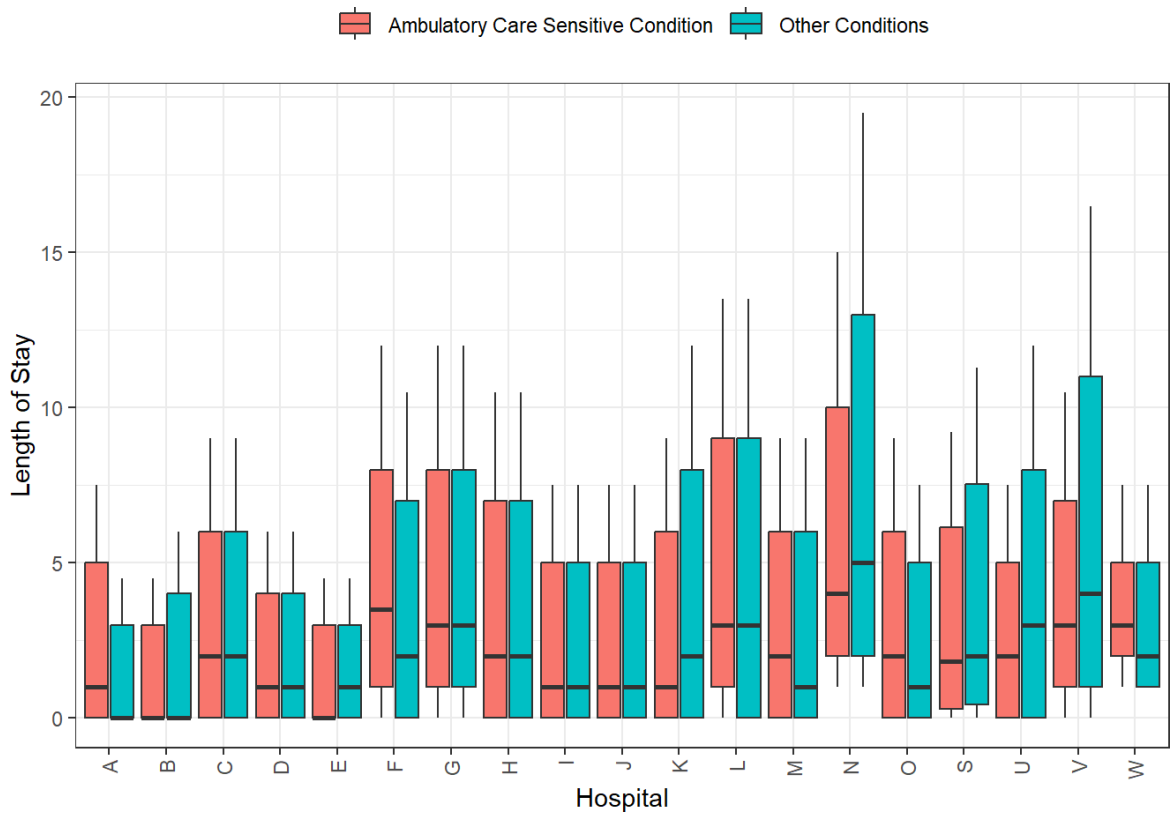
Figure APC6 suggests there is a higher proportion of short stays (<2 days) for non-ACSC (49.5% vs 45.8%) and a higher proportion of longer stays (2 days or over) for ACSC conditions (54.2% vs 50.5%).

Figure APC6: Length of stay in hospital



There is a considerable amount of variation between length of stay across the participating hospital sites. It is also not consistent as to whether ACSC have longer lengths of stays or non-ACSC. The median length of stay ranges from 0 - 4 days for ACSC conditions and 0 - 5 days for non - ACSC. (Figure APC7)

Figure APC7: variation in length of hospital stay by site and ACSC non-ACSC.



Discussion

Key use case findings

The study found there were high rates of Avoidable Admissions across all sites – averaging 40% of all admissions. We saw considerable variation between all our sites, indicating there are differences in how admissions are being managed which requires more research. Deprivation was a key factor observed equally in ACSC and non-ACSC groups, and there was a high proportion of patients attending ED with ACSC. Further research is needed to establish clearer criteria for potentially avoidable admissions and same day emergency care-eligible patients.

This case study also identified patients in the ECDS and APC datasets were older in the ACSC groups on average than the non-ACSC groups. The high proportions of patients attending ED with ACSC could signify a failure of care in pre-hospital settings, such as primary and community care.

High proportions of admissions from ED and other sources (including primary care) with ACSC may indicate that current Same Day Emergency Care (SDEC) services are not meeting the needs of those patients at highest risk of admission. Longer hospital stays for ACSC conditions may reflect the average age of these patients being higher. However, the fact that some short stay patients are not being captured amongst the ACSC group probably indicates that the definitions for SDEC patients are not sufficiently wide to capture these, generally younger, patients.

The approach was successful for aggregating real-world data for analysis of hospital admissions - some challenges were identified also which are highlighted below.

Key data capability and infrastructure learnings and challenges

Phase two of the regional linked health data programme has enabled the network to extend its reach across eight regions, all at varying levels digital maturity, data capability and TRE/SDE development. The programme has highlighted the value and opportunity in accessing regional health data, which often provides much of the granularity, depth and clinical context missing from national datasets. It also demonstrates the difficulty in developing a harmonised approach to data access, validation, processing and analysis when coordinating a multi-regional study with variation in data capability and linkage, coding, information governance and data access processes.

- **Complexity of local ethics and information governance processes**
  As demonstrated in Phase 1, navigating the varied local research and information governance frameworks and processes again contributed to significant delays in data approvals and access. For those regions which had REC Research Database approval (iCARE, DataLoch, PIONEER) the local approval process to access data was more streamlined as activity was being conducted within conditions of broader Research Database Approval, while other sites had to apply for access to

data on per project basis, which introduced greater administrative burden and significantly more delays to data access. This also contributed to the difficulty in the classification of this study, as either research or service evaluation.

- **Consensus on whether project was classified as research or service evaluation**
  Reaching a harmonised consensus across regions on the classification of the use cases as either research or service evaluation proved difficult. There was an equal split in the interpretation of the local row level analysis by the participating regions (see Table 1 and 2). There was a difference in how this type of project – with analysis on deidentified individual row level data undertaken locally and aggregated data only sent to lead site for analysis – was categorised by local information and research governance departments. Equally, there was uncertainty surrounding the research governance approach for the aggregated data analysis and whether this should be classified as research requiring ethical review by an NHS Research Ethics Committee (REC). Advice was sought via the Health Research Authority and sponsor for the project (University of Sheffield) who confirmed REC approval was not required but this process introduced significant delay. More clarity on how these types of projects are categorised are needed.

- **Complexity and variation in data access templates**
  Whilst the pathway to data access differs between regions, there are opportunities for alignment and standardisation. A good demonstration of this is in the templates used for Data Protection Impact Assessments (DPIAs) and Data Sharing Agreements (DSAs), which were required for Caldicott Guardian approval by several sites. Not surprisingly, the forms varied between regions but in some instances also differed within the same ICB. This was observed within South Yorkshire, where four distinct DPIAs were required to gain access to data from NHS Trusts in the region (namely Barnsley Hospital NHS Foundation Trust, Sheffield Teaching Hospitals NHS Foundation Trust, Rotherham NHS Foundation Trust and Doncaster and Bassetlaw Teaching Hospitals NHS Foundation Trust). This presented a significant administrative burden and resulted in a delay in accessing the necessary data. These delays could have been avoided were a standard template available, and there are opportunities to align with current initiatives such as the Pan UK Data Governance Steering Group led by HDR UK which is leading work to develop harmonised DSA which can be implemented across different regional TREs/SDEs. In addition, as the NHS England Data for R&D Programme is rolled out across England with 11 sub national SDEs, there are opportunities to harmonise and develop consistent standards for data access.

- **Data availability**
  Some regions have access to key data (such as primary care) as part of their existing IG and ethics approvals, whilst others do not. This was a limiting factor for the chest pain use case, which required data from primary data (but this was considered desirable, rather than an essential requirement). More widely however, access to linked regional primary care data within regional TREs/SDEs is a high priority and will be key for implementation of different use cases and is important in establishing population denominator for epidemiological research questions.

- **Challenges in data harmonisation**
  Challenges were identified in developing a single data analysis pipeline due to starting to work on the pipeline after the project had started - collaborative open-source development of

reproducible analytics pipelines (RAPs) can save significant time for all regional collaborators, but requires thorough planning from the outset. The lead site for use cases such as these should coordinate not only the protocol and documentation, but also the data validator, to prevent excessive workload for the aggregating (lead) site when data is submitted. In addition, the analytical team within some of the sites (e.g South Yorks), did not have any visibility of the data until it had been passed from their regional NHS Trusts to the research team, therefore it was not possible to gauge data quality (e.g., completeness) or to do any preparatory work on data validation or harmonisation. This led to delays for this site compared to other sites who had visibility of the data and were able to write validation scripts whilst waiting for governance approval. Finally, promotion of code sharing among all sites to facilitate efficient and consistent sub-analyses is important also- enabling the fostering an environment of open science, collaboration and continuous improvement.

- **Approach to data minimisation**
  Local protocols for Statistical Disclosure Control (SDC) was varied across regions, and proved difficult to reconcile, as demonstrated by the low number suppression threshold for each region. Without low number suppression it is impossible to adequately reduce the risks of re-identification in a complex aggregated dataset, however delays were introduced while harmonised protocol and approach for SDC across all sites was agreed.

Recommendations

Phase 1 of the Regional Linked Data Group listed key recommendations, identified through challenges experienced with the Phase 1 'driver' use case:

- **Standardisation and coding across regions**, with specific recommendation to identify and address the barriers that limit the use of the NHS-mandated SNOMED coding system at the point of care
- **Enabling streamlined, standardised and harmonised access to data** for approved researchers within regional TREs/SDEs via a national approval process modelled on current ethics approvals
- Regional TREs/SDEs to receive **seamless and streamlined access to the data arising from or pertinent to their local health system**s that is collated and curated nationally.
- **Leverage regional expertise embedded within each ICB** - with ICBs supported to train and support talented analysts from across the data ecosystems within their region.
- **Ongoing patient and public involvement and engagement** to enable transparency on how regional level data can used for research and health and care planning and public trust.

Phase 1 recommendations were as relevant for both Phase 2 'driver' use cases with challenges identified across aspects experienced and aligned to Phase 1 recommendations.

Phase 2 learnings and challenges have generated additional recommendations detailed below:

- **Regions which had Research Database Approval in place demonstrated that access to data for specific projects was more streamlined and rapid than those which had to apply for local ethics and IG approvals on a per project basis**. Recommend that regional TREs/SDE seek to obtain Research Database Approval, enabling a more streamlined approach to multi regional collaborative projects and earlier clarity around approved data uses.
- **Guidance is needed for local research and information governance offices to define whether these types of projects are categorised as service evaluation or research** which are sharing anonymised, aggregated data only across regions– for both use cases in Phase 2, local row level analysis was classed as service evaluation by some sites and research by others.
- **Implementation of harmonised and standardised Data Sharing Agreements (DSAs) and Data Protection Impact Assessment (DPIAs) templates to enable streamlined multi regional collaborations –** there is much complexity and variation in standard templates – such as data access forms, DPIA and DSAs, which not only varied between regions but also between hospitals sites within the same ICB.
- **Ensure development of Reproducible Analytical Pipelines (RAPs) for multi-regional projects are developed at the outset** and robust principles of open science are implemented to enable code sharing and reproducibility across regions and for possible future analyses using curated datasets.
- **Data availability** - some regions have access to key linked data (such as primary care) as part of their existing IG and ethics approvals, whilst others do not which can hamper collaboration and priority research and health and care planning questions to be addressed rapidly and at scale. Promote harmonisation of an agreed list of priority nationally and regionally collated specific data flows across multi-regional TREs/SDEs (e.g primary care, real time hospital admission discharge and transfer, secondary care laboratory and prescribing data)
- **Develop harmonised protocols across regions for Statistical Disclosure Control/ data minimisation** – in compliance with established guidance and good practice[7]

Next steps will include driving forward progress to implement approaches to standardisation and harmonisation across collaborators within the HDRUK Regional Linked Data Group, but also to share learnings and insights with national initiatives – such as the NHS England sub-national SDEs via the Data for R&D Programme and Scottish network of Regional Data Safe Havens. Scaling up of the Group to include more regions, (including exploring possibility of Wales and Northern Ireland joining), mapping onto and aligning with UK wide regional initiatives including HDR UK Regional Network will also be priorities for future work.

[7] https://ukdataservice.ac.uk/app/uploads/thf_datareport_aw_web.pdf

Appendix 1

Protocol Variation in acute admissions _fin