

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<u>http://bmjopen.bmj.com</u>).

If you have any questions on BMJ Open's open peer review process please email <u>info.bmjopen@bmj.com</u>

BMJ Open

The impact of longer patient travel distances and times on perioperative outcomes after complex revision knee surgery

Journal:	BMJ Open
Manuscript ID	bmjopen-2024-085201
Article Type:	Original research
Date Submitted by the Author:	15-Feb-2024
Complete List of Authors:	Matthews, Alexander; University of Exeter Faculty of Health and Life Sciences, Department of Clinical and Biomedical Sciences; Royal Devon and Exeter NHS Foundation Trust, Evans, Jonathan; University of Exeter Medical School, Health Services and Policy Research; Royal Devon and Exeter NHS Foundation Trust, Princess Elizabeth Orthopaedics Centre Evans, Jonathan; University of Exeter, Department of Public Health and Sports Sciences Lamb, Sarah; University of Exeter, College of Medicine and Health Price, Andrew; University of Oxford, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences Gray, William; NHS England and NHS Improvement London Briggs, Tim; NHS England and NHS Improvement London Toms, Andrew ; Princess Elizabeth Orthopaedic Centre
Keywords:	Knee < ORTHOPAEDIC & TRAUMA SURGERY, PUBLIC HEALTH, Health Services
Note: The following files were s PDF. You must view these files	ubmitted by the author for peer review, but cannot be converted to (e.g. movies) online.
TravelTimesCodeFinal.R	

SCHOLARONE[™] Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our <u>licence</u>.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which <u>Creative Commons</u> licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

terez oni

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Title

The impact of longer patient travel distances and times on perioperative outcomes after complex revision knee surgery

Names, Affiliations, and positions of all authors

Alex Matthews1,2,3, Jonathan P Evans2,3, Jonathan T Evans2,3, Sallie Lamb3 Andrew Price4,5, William Gray1, Tim Briggs1,6, Andrew Toms2,3

- 1. Getting It Right First Time programme, NHS England, London, UK
- 2. Royal Devon University Healthcare NHS Foundation Trust, Exeter, UK
- 3. University of Exeter, Exeter, UK
- 4. Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, UK
- 5. Nuffield Orthopaedic Centre, Oxford, UK
- 6. Royal National Orthopaedic Hospital, Stanmore, London, UK

Correspondence: Alexander Matthews, Getting It Right First Time programme, NHS England and NHS Improvement, Wellington House, Waterloo Road. London, UK, SE1 8UG; telephone: 07401932627 email: alexander.matthews1@nhs.net

Word count: [3316 words]

Short title: Patient Journey Distance and Time for Complex Revision Knee Surgery

Key words: social determinants of health, revision knee replacement, contextual factors in complex health interventions

Twitter handles:

Alex Matthews: @DrAlexMatthews

Tim WR Briggs: @ProfTimBriggs

WilliamKGray1: @WilliamKGray1

GIRFT programme: @NHSGIRFT

Structured Abstract (Word count suggested 250-300)

Objectives

Patients with problematic knee replacements requiring further surgery often have difficulties mobilising and increasingly rely on family support. Evolving practice in England aims to manage these patients in specialised centres with the intention of improving outcomes. This practice will result in longer travel distances and times in this frailer group of patients. We want to examine the types of distances and travel times patients can be expected to travel for complex orthopaedic surgery and to explore concerns of how these impact patient outcomes.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Design

Retrospective observational study from the Hospital Episode Statistics. Multivariable adjusted logistic regression modelling was used to compare the exposure variable with perioperative outcomes

Setting

Patients presenting to tertiary referral centres between 1st January 2016 to 31st December 2019. A tertiary referral centre was defined as a trust performing >70 revisions in the year prior.

Participants

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Adult patients undergoing revision total knee replacement procedures for aseptic reasons between 1st January 2016 to 31st December 2019.

Interventions

Patient level travel distance and time was calculated using the department of health Journey Time Statistics.

Main Outcome Measures

The primary outcome is the association of travel distance and time on emergency readmission within 30 days. Secondary outcomes will focus on mortality within 90 days and length of inpatient stay.

Results

1516 patients were treated at 16 tertiary referral centres for non-infected reasons. Patients in the longest driving distance group were expected to travel a median distance of 44.55 miles (IQR 35.90 to 56.30) with an expected median journey time of 66.3 minutes (IQR 57.9 to 80.5). Overall, 30-day readmission was not statistically associated with farther travel distances or driving times.

Conclusions

Patients were expected to travel up to hour for revision knee replacement surgery. There was no association between increasing travel distance and time on perioperative outcomes.

Summary Boxes

What is already known on this topic?

- A failed primary knee replacement is a life changing event often linked to reduced mobility and depression.
- Evolving practice in revision knee replacement surgery in England aims to treat these complex frail patients in super-specialised regional hospitals.
- Subsequently patients can expect to travel longer distances and times and it is unknown what affect these will have on patient outcomes.

What does this study add?

- Patient in the longest journey time category were expected to travel over an hour at peak driving times.
- Patient's travelling farther for revision knee replacement surgery did not demonstrate any statistically worse perioperative clinical outcomes.
- This information is of utility to surgical providers and commissioners of healthcare services and can inform patient-led decision-making surrounding travelling for complex revision knee replacement surgery.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Introduction

Primary knee replacement is a successful procedure that improves quality of life for the majority of patients.(1) However, at 10 years following a primary knee replacement, about 3.5% of patients will have undergone a revision surgery.(2) The majority of these procedures are carried out due to infection or polyethylene wear of the implant.(3) A failed primary knee replacement represents a life changing transition point where individuals are likely to suffer from pain, reduced mobility as well as dependency on family members.(4) Patients often face multi-step surgery with longer hospital length of stays and higher complication rates.(5, 6)

The orthopaedic GIRFT (Getting It Right First Time) programme was launched in 2012 following the publication of the Orthopaedic National Report.(7) A key recommendation was the centralisation of complex orthopaedic surgery, including revision knee surgery, to specialist centres with the aim of improved patient outcomes. Consequently, revision total knee replacement (rTKR) surgery in the England has evolved into a regional network service model. .(8) In doing so, all hospitals performing rTKRs form a network in the respective regions. Less specialist hospitals defined by lower annual case volume thresholds are encouraged to discuss and sometimes refer their caseload to more specialist centres. Several studies based on large revision hip and knee registries has suggested this model carries a lower failure rate defined by the need for further revision surgery.(9-11) Early evidence has suggested reduced early failure rates through the adoption of revision knee networks.(12)

However, this approach to managing patients is inevitably associated with increasing travel distances between some patient's homes and their treating hospital. Expected distances are important to explore, particularly as these patients struggle with mobility, may be unable to drive and may be more reliant on family members.(4) Furthermore, greater travel distances have been associated with higher readmission rates and higher mortality rates following complex vascular surgery.(13) There is also concern that patients required to travel greater distances are more likely to be re-admitted to a different hospital resulting in clinical decisions that do not

incorporate the primary surgeon and potentially alter outcomes.(14) Subsequently the aims of this paper is to examine if the same association with longer patient travel distance and perioperative outcomes exists following complex orthopaedic surgery with a focus on revision knee replacement surgery performed in high volume tertiary referral centres.

Methods

Design

This study is a retrospective data analysis of observational data from the Hospital Episode Statistics (HES) and Office for National Statistics (ONS) databases. HES data is collected by NHS England for all patients treated at NHS hospitals in England and those treated at private hospitals where treatment was funded by the NHS. This study complies with the recommended reporting guidelines when using HES data(15) and the Strengthening of Reporting of Observational studies in Epidemiology (STROBE) guidelines.(16)

The analysis and presentation of data follows current NHS England guidance for the use of HES data for research purposes(17) and is anonymised to the level required by ISB1523 Anonymisation Standard for Publishing Health and Social Care Data.(18) The HES data were linked at a patient level to data from the ONS on deaths and date of death, which allowed the identification of patients who had died after their surgery. Linkage was achieved using a unique patient identifier using a previously validated methodology.⁽¹⁹⁾

Patient travel distances were calculated using the Journey Time Statistics reference document produced by the UK Department of Transport which modelled theoretical journey times between known Lower Layer Super Output Areas (LSOA) of residence and NHS hospital sites.(20) The Journey Time Statistics document is available in the supplementary material section.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Population

An rTKR procedure was defined as a permanent removal or exchange of knee arthroplasty components. This includes a revision of a total knee replacement and a conversion of a unicondylar knee replacement to a total knee replacement. Secondary patellar resurfacing was not included as this represents a simple revision procedure, one that can be carried out in most nonspecialised hospitals. All patients aged \geq 18 years who underwent a rTKR in a high-volume trust between 1st January 2016 to 31st December 2019 were included in the study population. A high-volume trust was classified as a centre performing > 70 revisions per year. This revision volume threshold for classification represents those proposed by the British Association for Surgeons of the Knee (BASK) Revision Knee Working Group. (21) These are more likely to represent tertiary referral centres where the stratification of more complex work will take place. Annual case volume at each trust was defined as the number of revision cases conducted in the year prior to the index procedure. This measure was preferred over a simple calculation of average annual volume as it accounts for recent experience at the point of surgery. The Office for Population Censuses and Surveys' Classification of Interventions and Procedures version 4 (OPCS-4) codes used to identify rTKR procedures are detailed in Supplementary material S1. Where the procedure laterality was not specified, patients were excluded. The flow of patients, with numbers excluded at each point, is summarised in Supplementary material S2. To manage population heterogeneity, data were extracted for the period 1st April 2011 to 31st December 2019 and only the first revision for a specific side of the body record in this time period included. (22) Thus, any early revisions on the same side of the body in the four years and nine months preceding the start of the study period were identified and these patients excluded from the study. This aims to exclude the early revision knee replacement failures which have been shown to represent catastrophic failures potentially skewing our results. (22)We excluded revisions for infection as these represent a more variable patient group with a different complication profile (23) and this is further discussed in our study limitations.

Exposure variable

In the analysis straight line travel distance was calculated between a patient's LSOA and the postal codes for the treating hospitals. LSOAs are determined by the Office

for National Statistics and are designed for the reporting of small area statistics. Public transport and highways data for England was used to create theoretical journey distances and times from origins to destinations. The resulting travel distances and/or times for each patient were divided into quintiles following a recently reported methodology.(13) Sensitivity analyses were performed using travel distances by road and peak driving times to account for variation in travel infrastructure between rural and urban areas and to attribute more meaningful results for patients. Peak driving times were calculated by using average traffic speeds for between 7am and 10am.

Outcomes

The primary outcome was emergency readmission within 30 days of discharge from the index surgical hospital. Secondary outcomes included 90-day mortality, and hospital length of stay (LOS) above the median. The LOS outcome was dichotomised into above median or below median LOS of five days.

Statistical Analyses

Data was extracted from a secure, encrypted server controlled by NHS Digital. Data were analysed within a secure, encrypted environment using standard statistical software: R Studio version 2023.09.1+494 (Boston, Massachusetts, USA). The R code and packages used are included in **Supplementary material S3**

Crude comparisons of baseline categorical characteristics and travel distance proximity were calculated. A This data were categorical in nature and summarised as frequency and percentage. In primary analysis a logistic multivariable regression model was constructed to evaluate associations between travel distance quintiles and 30 day readmission, with adjustment for the covariates listed above. The first (shortest) travel distance quintile was used as the reference in all models.

Age, sex , comorbidities and characteristics of initial presentation were included in the logistic regression model. These variables have been shown to influence the risk of complications after R-TKA and therefore represent known confounders.(9, 10, 23) This also included data on economic deprivation measured using the Index of Multiple Deprivation (IMD).(24) The IMD gives the LSOA where the patient lives a score based on a range of measures of deprivation. IMD was categorised into quintiles, based on all-England data, for analysis. A spearman's rank correlation was performed to investigate the relationship between IMD score and travel distances. Health co-morbidity was quantified using the Hospital Frailty Risk Score (HFRS). HFRS identifies frailty based on the occurrence of any of 109 International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10) codes used during any hospital admissions in the two years prior to, and for, the index admission.

All secondary outcomes were binary and analysed using the same multivariable logistic regression. Multicollinearity was assessed with reference to variance inflation factor and Shapiro-Wilcox test of normality. Model fit was assessed with reference to the pseudo R² values.

Results

Demographic characteristics and co-morbidities

The 1,516 patients in the overall study population, were divided into quintiles of travel distance of 303 or 304 patients each. The median straight line travel distance for quintile one was 2.3 miles (IQR 1.3 to 3.1). For the fifth travel quintile, median distance was 33.5 miles (IQR 25.5 to 41.1). Baseline co-morbidities and demographic characteristics were broadly similar among the travel distance quintiles (**Table 1**). Travel distance was not strongly correlated with age or social deprivation (**Figure 1**)

Association between travel distance and readmission, mortality and extended hospital stay

Overall, 111 patients were readmitted within 30 days. Crude comparisons of proportions readmitted within 30 days for each travel distance quintiles revealed a higher rate of readmission for the second travel quintile. In multivariable adjusted logistic regression, there was no statistically significant association between travel distance and readmission within 30 days (**Table 2**). Odds for 30-day readmission was 1.44 (95% CI 0.71 to 2.96, P 0.32) for Q5 compared with Q1. Increased travel distance was not associated with a significant change in the odds of death within 90

days (OR for Q5 vs Q1, 1.46 (95% CI 0.49 to 4.53, P 0.682)). Travel distance quintile was not associated with prolonged length of hospital stay related to the index surgery after multivariable adjustment (OR for Q5 vs Q1, 0.96 (95% CI 0.67 to 1.39, P 0.84)).

Real world travel distance and outcomes

The above results used straight line travel distance between patient' LSOA and treating hospital. A sensitivity analysis using actual patient travel distances using the shortest possible road route was performed **(Table 3)**. The median driving distance by the shortest possible road route for the closest quintile was 3.40 miles (IQR 2.00 to 4.40). The furthest quintile median driving distance was 44.55 miles (IQR 35.90 to 56.30) This analysis showed no association between driving distance and all perioperative outcomes. The OR for readmission within 30 days in Q5 vs Q1 was 1.16(0.56-2.41, p value = 0.68).

Journey drive times and outcomes

A further sensitivity analysis using driving times was calculated (Table 4). The median drive time for quintile one was 12.6 minutes (IQR 8.7 to 15.3). For the fifth time quintile the median was 66.3 minutes (IQR 57.9 to 80.5). No statistical association was found between drive time and perioperative outcomes. The OR for readmission within 30 days in Q5 vs Q1 was 0.92 (0.45 - 1.85, p value = 0.81)

Discussion

Statement of principal findings

We present a multi-hospital site retrospective analysis of patients undergoing revision knee replacement surgery at tertiary referral centres in England. In this analysis of 1,516 patients undergoing aseptic revision knee replacement surgery, we did not observe an association between distance and time travelled for revision surgery and readmission within 30 days. Patients in the longest driving time category were expected to travel for a median time of more than one hour.

BMJ Open

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Strengths and weaknesses of the study

The findings of this study should be interpreted in view of several limitations. Firstly, this analysis used observational data from a large administrative dataset covering all NHS-funded procedures conducted in England. As with all datasets we are limited in the amount of detail provided regarding presentation. We chose to categorise a highvolume centre by trust to accurately capture surgical experience. All NHS hospitals in England are run by hospital trusts which typically involve between one and four hospitals within a catchment area standardising their practice. It is common practice for specialist orthopaedic surgeons to move between these sites delivering the same procedures. Our study involved 35 hospital sites run by 16 trusts. We acknowledge this is a weakness of our study as this may not be representative of all trusts. The indication for revision coded as mechanical complication encompasses several common indications such as aseptic loosening, instability and malalignment. Reassuringly these indications have similar length of stay, and perioperative outcomes.(23) Differences exist in their re-revision rate, however this was not an outcome of focus in our study. It is likely that the complexity of the surgery undertaken may vary within the different indications for revision. Evidence suggests that operative surgical time is related to increased length of stay in aseptic revision knee replacement.(25). There is a lack of granular data for revisions due to infection and therefore we excluded this patient group as some readmissions for this patient group may represent planned readmissions. There is also a lack of granular clinical data using HES for each readmission, therefore we cannot ascertain precise reasons for readmissions, but we assume are related to a post-surgical complication. Clinical coding practice within HES is known to vary across trusts. (26) As an example, some trusts may be more consistent in coding comorbidities, and this may have created some bias. However, this is unlikely to vary systematically with travel distances and so significantly bias our findings. We acknowledge the relatively short travel distances in this population compared to examples from the United States as such the results of this study may not be generalisable to larger geographical areas or less mature healthcare systems. However, the upper guintile in our study represents a substantial journey distance and time for our patient cohort where poor mobility is a

BMJ Open

significant factor affecting their care. This analysis does not consider journey times of those who may not have access to a car and instead chose to take public transport.

Strengths and weaknesses in relation to other studies, discussing important differences in results

This is the first study to analyse the potential impact of patient travel distances on patients receiving complex orthopaedic surgery. The findings that longer travel distances are not associated with inferior outcomes is an important part of the evaluation of the assumptions and context behind the establishment of revision knee networks.(27) This study has shown that concerns of introducing a network in larger geographical regions, for example in Scotland where longer patient travel distances and times are common, may be less important.(28) This is particularly useful as regions explore the geography of their revision networks and during summative outcome assessment of this complex health intervention.(29)

It may be seen as surprising that no association between travel distance and prolonged length of hospital stay was identified. An expectation exists of increasing difficulties being encountered with the discharge of patients living greater distances from their treating hospital, which has been observed in patients following elective pancreatic surgery.(30) This is also an observation seen in patients being treated in specialist vascular centres in the United States which led to the recommendation of additional care coordination and follow up efforts. However, the geography of the population in these studies was much larger with significantly longer travel distances.

We did not observe a strong correlation between social deprivation status and age of the patient with longer travel distances. It is reassuring that access to treatment for older patients and those from poor socioeconomic backgrounds is unaffected by travel distance. However, there may be patients who refused to travel to a specialist centre and opted for treatment at their local centre.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Meaning of the study: possible explanations and implications for clinicians and policymakers

The organisation and delivery of revision knee services in England has recently undergone a substantial change and now such services are provided around regional networks of care. This promises substantial advantages to the increasing number of patients with problematic knee replacements in our ageing population who will benefit from regional expertise.(8) However, it is unknown the impact of patients residing farther from tertiary referral centres, particularly rural patients who may encounter additional difficulties associated with greater travel distance. A recent study following the outcomes of aortic surgery found that longer travel distances are associated with inferior perioperative outcomes(13). Similar associations have been found in postoperative colorectal surgery patients (31). As such our results are reassuring to policy makers and clinicians.

Unanswered questions and future research

There is a scarcity of evidence evaluating the patient perception of complex health interventions such as network models of care. Recent work by Kugler et al has demonstrated the willingness of patients to travel further for better outcomes in the context of total knee replacement surgery. (32) Nevertheless, patient perceptions of travelling further for their treatment should be a focus for future research in the context of revision knee patients, particularly as this is one of the top ten research priorities identified by the James Lind Alliance priority setting partnership.(33)

Conclusion

We did not observe an association in our study population between 30-day readmission rates and increasing travel distances or times between a patient's home and their treating hospital in revision knee replacement. This paper is the first to

explore the relationship between travel distance and complex orthopaedic surgery and informs some concerns regarding the creation of a centralised revision knee network. This information is of utility to surgical providers and commissioners of healthcare services. Furthermore, it can inform patient-led decision making and the exploration of perceptions surrounding travelling for complex surgery. Although this is the first assessment in complex orthopaedic surgery, a prospective analysis will be undertaken as part of the ongoing auditing of revision knee networks in England.

Supplementary material, Figures, Tables and Files

Supplementary material S1 – OPCS-4 code criteria used for Hospital Episode Statistics data extraction

See separate file named supplementary material

Supplementary material S2 – Flow of patient inclusion/exclusions

-See attached file named Supplementary Material

Supplementary material S3 - R Code

See attached file named Supplementary Material

Figure 1 – Box plot showing association of social deprivation and age on travel distance quintile. Spearman's rank correlation investigating relationship between these factors with travel distance.

-See attached file names tables and figures

Table 1 – Baseline demographics and clinical characteristics and raw perioperative outcomes for patients by travel distances quintile

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Number of Patients	304	303	303	303	303
Deprivation Quintile					
1(Most Deprived)	54 (18%)	87 (29%)	50 (17%)	58 (19%)	57 (19%)
2	48 (16%)	61 (20%)	68 (22%)	58 (19%)	66 (22%)
3	57 (19%)	65 (21%)	63 (21%)	63 (21%)	56 (18%)
4	55 (18%)	43 (14%)	49 (16%)	80 (26%)	75 (25%) 🦳
5 (Least Deprived)	90 (30%)	47 (16%)	73 (24%)	44 (15%)	49 (16%)
Sex					
Male	121 (40%)	133 (44%)	135 (45%)	124 (41%)	141 (47%)
Age in years					
16-59	52 (17%)	42 (14%)	56 (18%)	57 (19%)	68 (22%)
60-64	46 (15%)	29 (10%)	35 (12%)	38 (13%)	47 (16%)
65-69	53 (17%)	58 (19%)	56 (18%)	52 (17%)	44 (15%)
70-74	45 (15%)	56 (18%)	53 (17%)	59 (19%)	49 (16%)
75-79	46 (15%)	43 (14%)	47 (16%)	44 (15%)	47 (16%)
>=80	62 (20%)	75 (25%)	56 (18%)	53 (17%)	48 (16%)
Diagnosis					
Mechanical Complication	172 (57%)	198 (65%)	208 (69%)	212 (70%)	192 (63%)
Fracture	26 (9%)	29 (10%)	18 (6%)	10 (3%)	31 (10%)
Progressive OA	39 (13%)	32 (11%)	24 (8%)	24 (8%)	18 (6%)
Hospital Frailty Risk Score					
None	158 (52%)	115 (38%)	149 (49%)	156 (51%)	158 (52%)
Mild	100 (33%)	123 (41%)	106 (35%)	100 (33%)	99 (33%)
Moderate	38 (13%)	56 (18%)	40 (13%)	43 (14%)	36 (12%)

Severe

Volume 0-4

Volume 5-9

Volume 10-14

Volume 15-19

Volume 20-24

Volume >=25

90 Day Mortality

Annual Surgeon Volume

Perioperative Outcomes Readmission within 30 days

Prolonged Length of Stay

4 (1%)

30 (10%)

41 (14%)

63 (21%)

47 (16%)

62 (20%)

60 (20%)

17(6%)

7(2%)

130(43%)

10 (3%)

24 (8%)

35 (12%)

58 (19%)

53 (17%)

47 (16%)

86 (28%)

20(7%)

10(3%)

132(44%)

8 (3%)

31 (10%)

44 (15%)

55 (18%)

44 (15%)

74 (24%)

55 (18%)

23(8%)

10(3%)

135(45%)

8 (3%)

30 (10%)

43 (14%)

89 (29%)

49 (16%)

48 (16%)

45 (15%)

15(5%)

8(3%)

135(44%)

9 (3%)

39 (13%)

46 (15%)

72 (24%)

56 (18%)

64 (21%)

26 (9%)

36(12%)

134(44%)

15(5%)

1	
2	
3	
4	
5	
6	
7	
, 8	
٥ ۵	
1	n
1	1
1	י ר
1	2 2
1	3 4
1	4
1	5
1	6
1	/
1	8
1	9
2	0
2	1
2	2
2	3
2	4
2	5
2	6
2	7
2	8
2	9
3	0
~	

54

	1
For peer review only - http://bmjop	en.bmj.com/site/about/guidelines.xhtml

Table 2 – Multivariable Logistic Regression showing Odds Ratios for primary and secondary outcomes by straight line travel quintile

	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Readmission with 30	2.23 (95% CI 1.19 to	1.55 (95% CI 0.79 to	1.06 (95% CI 0.51 to	1.44 (95% CI 0.71 to
days	4.37),p=0.01	3.13),p=0.21	2.23),p=0.87	2.96),p=0.32
90 Day Mortality	1.79 (95% CI 0.66 to	1.55 (95% CI 0.52 to	1.72 (95% CI 0.53 to	1.46 (95% CI 0.49 to
	5.17),p=0.27	4.70),p=0.43	5.64),p=0.36	4.53),p=0.50
Prolonged Length of	0.90 (95% CI 0.62 to	1.02 (95% Cl 0.71 to	0.99 (95% CI 0.69 to	0.96 (95% CI 0.67 to
stay	1.30),p=0.57	1.46),p=0.91	1.412),p=0.95	1.39),p=0.84

Table 3 – Sensitivity analysis exploring road travel distance quintile and primary/secondary outcomes

	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Readmission with	2.11 (95% CI 1.14 to	1.41 (95% CI 0.71 to 2.84),p=0.33	1.44 (95% CI 0.73 to	1.16 (95% CI 0.56 to
30 days	4.06),p=0.02		2.90),p=0.29	2.4),p=0.68
90 Day Mortality	1.52 (95% CI 0.58 to	1.18 (95% CI 0.39 to	2.42 (95% CI 0.83 to	0.83 (95% CI 0.25 to
	4.21),p=0.40	3.56),p=0.77	7.27),p=0.11	2.64),p=0.75
Prolonged Length of stay	0.69 (95% CI 0.48 to 0.99),p=0.04	0.94 (95% CI 0.66 to 1.35),p=0.74	0.86 (95% CI 0.60 to 1.23),p=0.41	0.95 (95% CI 0.66 to 1.36),p=0.77

Table 4- Sensitivity analysis exploring driving time quintile and primary/secondary outcomes

	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Readmission with 30 days	1.19 (95% CI 0.63 to 2.25),p=0.59	1.23 (95% CI 0.65 to 2.35),p=0.52	1.53 (95% CI 0.82 to 2.89),p=0.18	0.92 (95% CI 0.45 to 1.85),p=0.81
90 Day Mortality	2.29 (95% CI 0.84 to 6.76),p=0.12	1.52 (95% CI 0.50 to 4.74),p=0.46	2.63 (95% CI 0.90 to 8.16),p=0.08	0.94 (95% CI 0.26 to 3.23),p=0.91
Prolonged Length of stay	0.82 (95% CI 0.57 to 1.52),p=0.27	0.80 (95% CI 0.56 to 1.15),p=0.23	1.27 (95% CI 0.89 to 1.81),p=0.19	1.01 (95% CI 0.70 to 1.44),p=0.97
			22	

Public and Patient Involvement statement

The study's chief investigator (AT) led the James Lind Alliance 'Revision Knee Replacement' priority setting partnership. This group of patients, carers and health care professionals identified the need to investigate the best way of organising revision knee replacement surgery to improve patient outcomes as one of their top 10 research questions. Patients were therefore directly involved in the development of the study's aims and objectives. The results of the study will be disseminated to the members of this group prior to publication.

Copyright/licence for publication Statement

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence.

Competing Interests

All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

Transparency Declaration

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Ethical Approval

The analysis and presentation of data follows current NHS England guidance for the use of HES data for research purposes and is anonymised to the level required by ISB1523 Anonymisation Standard for Publishing Health and Social Care Data. Ethical approval was not required.

Funding

No funding was obtained to carry out this study

Data Sharing

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

STROBE Statement—Checklist of items that should be included in reports of *cohort studies*

	Item No	Recommendation	Page No S
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract(b) Provide in the abstract an informative and balanced summary of what was	2 2,3
		done and what was found	

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

Background/ratio	nale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives		3	State specific objectives, including any prespecified hypotheses	
Mathada				
Study design		4	Present key elements of study design early in the naner	
Setting		5	Describe the setting locations and relevant dates including periods of	
Setting		5	recruitment, exposure, follow-up, and data collection	
Participants		6	(a) Give the eligibility criteria, and the sources and methods of selection of	
1			participants. Describe methods of follow-up	
			(b) For matched studies, give matching criteria and number of exposed and	
			unexposed	
Variables		7	Clearly define all outcomes, exposures, predictors, potential confounders, and	
			effect modifiers. Give diagnostic criteria, if applicable	
Data sources/		8*	For each variable of interest, give sources of data and details of methods of	
measurement			assessment (measurement). Describe comparability of assessment methods if	
D'		0	there is more than one group	_
Blas		9	Describe any efforts to address potential sources of bias	
Ouantitativa varia	blac	10	Explain now me study size was arrived at	+
Quantitative varia	iores	11	Explain now quantitative variables were nanoled in the analyses. If applicable,	
Statistical method	10	12	(a) Describe all statistical methods, including those used to control for	+
Statistical method	10	12	confounding	
			(b) Describe any methods used to examine subgroups and interactions	
			(c) Explain how missing data were addressed	
			(d) If applicable, explain how loss to follow-up was addressed	
			(e) Describe any sensitivity analyses	
Results		1.2*	(a) \mathbf{D} and \mathbf{D} (b) \mathbf{L} (c) \mathbf{L}	_
Participants		13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially	
			completing follow up and analyzed	
			(b) Cive reasons for non-norticipation at each stage	
			(a) Consider use of a flow diagram	
Descriptive data		1/1*	(a) Give characteristics of study participants (eq demographic clinical social)	+
Descriptive data		14	and information on exposures and potential confounders	
			(b) Indicate number of participants with missing data for each variable of interest	
			(c) Summarise follow-up time (eg. average and total amount)	
Outcome data		15*	Report numbers of outcome events or summary measures over time	╈
		10		
N. 1.	16	$() \cap$		0
iviain results	10	(a) Give i	manufusion estimates and, in applicable, confounder-adjusted estimates and their (ag. 05% confidence interval). Make clear which confounders were adjusted for	ð
		and why	(cg, 5570 confidence incrvar). Wake creat which confounders were adjusted for hey were included	
		(b) Report	t category boundaries when continuous variables were categorized	8
		(c) If role	vant consider translating estimates of relative risk into absolute risk for a	0 n/
		meaninof	ul time neriod	11/
Other analyses	17	Report of	her analyses done—eg analyses of subgroups and interactions, and sensitivity	8.9
unur 5000	- '	analyses		-,
Diagonation		<i></i>		
Discussion	10	Cummerati	a have regulte with reference to study chiesting	0
Limitations	18	Diaguage 1	mitations of the study, taking into account sources of notantial bios or	9 10
Limitations	19	improvisi	minations of the study, taking into account sources of potential bias or	10
Interpretation	20	Cive case	Discuss boin direction and magnitude of any potential blas	0
interpretation	20	Give a ca	ty of analyses, results from similar studies, and other relevant avidence.	У,
Generalisability	21	Discuse +1	e generalisability (external validity) of the study results	10
Seneralisability	<u>~1</u>	130035 ll	to Seneralisaonity (external valuaty) of the study results	1(
Other informatio	on	~ .		
Funding	22	Give the s	source of funding and the role of the funders for the present study and, if	23
		1. 11	. A set the second stand stand second big by the second sector by the terms of	

References

1. Price AJ, Alvand A, Troelsen A, Katz JN, Hooper G, Gray A, et al. Knee replacement. Lancet. 2018;392(10158):1672-82.

2. Sabah SA, Hedge EA, Abram SGF, Alvand A, Price AJ, Hopewell S. Patient-reported outcome measures following revision knee replacement: a review of PROM instrument utilisation and measurement properties using the COSMIN checklist. BMJ Open. 2021;11(10):e046169.

3. Sabah S, Knight R, Alvand A, Murray D, Petrou S, Beard D, Price A. No exponential rise in revision knee replacement surgery over the past 15 years: an analysis from the National Joint Registry. Osteoarthritis and Cartilage. 2022;30(12):1670-9.

4. Rowland T, Lindberg-Larsen M, Santy-Tomlinson J, Jensen CM. A qualitative study of patients' experiences before, during and after surgical treatment for periprosthetic knee joint infection; "I assumed it had to be like that ... ". International Journal of Orthopaedic and Trauma Nursing. 2023;48:100992.

5. Garbarino LJ, Gold PA, Sodhi N, Anis HK, Ehiorobo JO, Boraiah S, et al. The effect of operative time on in-hospital length of stay in revision total knee arthroplasty. Ann Transl Med. 2019;7(4):66.

6. Liodakis E, Bergeron SG, Zukor DJ, Huk OL, Epure LM, Antoniou J. Perioperative Complications and Length of Stay After Revision Total Hip and Knee Arthroplasties: An Analysis of the NSQIP Database. The Journal of Arthroplasty. 2015;30(11):1868-71.

7. Briggs TWR. A National Review of Adult Elective Orthopaedic Services in England. In: Time GIRF, editor. 2015.

8. Kalson NS, Mathews JA, Phillips JRA, Baker PN, Price AJ, Toms AD, et al. Revision knee replacement surgery in the NHS: A BASK surgical practice guideline. The Knee. 2021;29:353-64.

9. Yapp LZ, Walmsley PJ, Moran M, Clarke JV, Simpson AHRW, Scott CEH. The effect of hospital case volume on re-revision following revision total knee arthroplasty. The Bone & Joint Journal. 2021;103-B(4):602-9.

10. Halder AM, Gehrke T, Günster C, Heller KD, Leicht H, Malzahn J, et al. Low Hospital Volume Increases Re-Revision Rate Following Aseptic Revision Total Knee Arthroplasty: An Analysis of 23,644 Cases. J Arthroplasty. 2020;35(4):1054-9.

11. Jeschke E, Gehrke T, Günster C, Heller KD, Leicht H, Malzahn J, et al. Low Hospital Volume Increases Revision Rate and Mortality Following Revision Total Hip Arthroplasty: An Analysis of 17,773 Cases. J Arthroplasty. 2019;34(9):2045-50.

12. Bloch BV, Matar HE, Berber R, Gray WK, Briggs TWR, James PJ, Manktelow ARJ. The impact of a revision arthroplasty network on patient outcomes. Bone Joint J. 2023;105-b(6):641-8.

13. Feldman ZM, Zheng X, Mao J, Sumpio BJ, Mohebali J, Chang DC, et al. Longer patient travel distance is associated with increased non-index readmission after complex aortic surgery. Journal of Vascular Surgery. 2023;77(6):1607-17.e7.

BMJ Open

14. Saunders RS, Fernandes-Taylor S, Kind AJ, Engelbert TL, Greenberg CC, Smith MA, et al. Rehospitalization to primary versus different facilities following abdominal aortic aneurysm repair. J Vasc Surg. 2014;59(6):1502-10, 10.e1-2.

 15. Sinha S, Peach G, Poloniecki JD, Thompson MM, Holt PJ. Studies using English administrative data (Hospital Episode Statistics) to assess health-care outcomes--systematic review and recommendations for reporting. Eur J Public Health. 2013;23(1):86-92.

16. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344-9.

NHS Digital. Hospital Episode Statistics(HES) Analysis Guide. London, UK: NHS Digital;
2018.

18. Information Standard Board for Health and Social Care. Anonymisation Standard for Publishing Health and Social Care Data Specification. London, UK: NHS Digital; 2013.

19. NHS Digital. Delivering linked datasets to support health and care delivery and research. London, UK: NHS Digital; 2018.

20. Department for Transport. Journey Time Statistics, notes and definitions. London, UK: Department for Transport; 2019.

21. Kalson NS, Mathews JA, Miles J, Bloch BV, Price AJ, Phillips JRA, et al. Provision of revision knee surgery and calculation of the effect of a network service reconfiguration: An analysis from the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. Knee. 2020;27(5):1593-600.

22. Deere K, Whitehouse MR, Kunutsor SK, Sayers A, Price AJ, Mason J, Blom AW. How long do revised and multiply revised knee replacements last? An analysis of the National Joint Registry. Lancet Rheumatol. 2021;3(6):e438-e46.

23. Sabah SA, Knight R, Alvand A, Palmer AJR, Middleton R, Abram SGF, et al. Patient-Relevant Outcomes Following First Revision Total Knee Arthroplasty, by Diagnosis: An Analysis of Implant Survivorship, Mortality, Serious Medical Complications, and Patient-Reported Outcome Measures Utilizing the National Joint Registry Data Set. JBJS. 2023;105(20):1611-21.

24. Kamath CC, O'Byrne TJ, Lewallen DG, Berry DJ, Maradit Kremers H. Neighborhood-Level Socioeconomic Deprivation, Rurality, and Long-Term Outcomes of Patients Undergoing Total Joint Arthroplasty: Analysis from a Large, Tertiary Care Hospital. Mayo Clin Proc Innov Qual Outcomes. 2022;6(4):337-46.

25. Lindberg-Larsen M, Jørgensen CC, Bæk Hansen T, Solgaard S, Odgaard A, Kehlet H. Re-admissions, re-operations and length of stay in hospital after aseptic revision knee replacement in Denmark: a two-year nationwide study. Bone Joint J. 2014;96-b(12):1649-56.

26. Hardy F, Heyl J, Tucker K, Hopper A, Marchã MJ, Briggs TWR, et al. Data consistency in the English Hospital Episodes Statistics database. BMJ Health Care Inform. 2022;29(1).

27. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021;374:n2061.

28. Farrow L, Kennedy JW, Yapp L, Harding T, Walmsley P. Provision of revision knee arthroplasty services across Scotland: A national audit. Knee. 2023;42:312-9.

29. Paterson C, Baarts C, Launsø L, Verhoef MJ. Evaluating complex health interventions: a critical analysis of the 'outcomes' concept. BMC Complement Altern Med. 2009;9:18.

30. Jackson KL, Glasgow RE, Mone MC, Sheng X, Mulvihill SJ, Scaife CL. Does travel distance influence length of stay in elective pancreatic surgery? HPB (Oxford). 2014;16(6):543-9.

31. Cloyd JM, Huang L, Ma Y, Rhoads KF. Predictors of readmission to non-index hospitals after colorectal surgery. The American Journal of Surgery. 2017;213(1):18-23.

32. Kugler CM, De Santis KK, Rombey T, Goossen K, Breuing J, Könsgen N, et al. Perspective of potential patients on the hospital volume-outcome relationship and the minimum volume threshold for total knee arthroplasty: a qualitative focus group and interview study. BMC Health Serv Res. 2021;21(1):633.

alson hroplasty. 33. Mathews JA, Kalson NS, Tarrant PM, Toms AD. Top ten research priorities for problematic knee arthroplasty. Bone Joint J. 2020;102-b(9):1176-82.

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Tables and Figures

Figure 1 – Box plot showing association of social deprivation and age on travel distance quintile. Spearman's rank correlation investigating relationship between these factors with travel distance.



Spearman's Rank Correlation

Age and Travel Distance = rho -0.08 pvalue = 0.00126 (very weak correlation as age increases travel distance decreases)

Social Deprivation and Travel distance = rho -0.01 pvalue = 0.6

Table 1 – Baseline demographics and clinical characteristics and rawperioperative outcomes for patients by travel distances quintile

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Number of Patients	304	303	303	303	303
Deprivation Quintile					
1(Most Deprived)	54 (18%)	87 (29%)	50 (17%)	58 (19%)	57 (19%)
2	48 (16%)	61 (20%)	68 (22%)	58 (19%)	66 (22%)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

	a
1	
2	
3	
4	
5	
6	
7	
8	
9	
1	0
1	1
1	2
1	3
1	4
1	5
1	6
1	7
1	8
1	9
2	0
2	1
2	2
2	3
2	4
2	5
2	6
2	7
2	8
2	9
3	0
3	1
3	2
3	3
3	4
3	5
3	6
_	-

3	57 (19%)	65 (21%)	63 (21%)	63 (21%)	56 (18%)
4	55 (18%)	43 (14%)	49 (16%)	80 (26%)	75 (25%)
5 (Least Deprived)	90 (30%)	47 (16%)	73 (24%)	44 (15%)	49 (16%)
Sex	121 (10%)	122 (11%)	125 (15%)	174 (41%)	111 (17%)
	121 (4078)	133 (4478)	133 (4376)	124 (4170)	141 (4770)
16-59	52 (17%)	42 (14%)	56 (18%)	57 (19%)	68 (22%)
60-64	46 (15%)	29 (10%)	35 (12%)	38 (13%)	47 (16%)
65-69	53 (17%)	58 (19%)	56 (18%)	52 (17%)	44 (15%)
70-74	45 (15%)	56 (18%)	53 (17%)	59 (19%)	49 (16%)
75-79	46 (15%)	43 (14%)	47 (16%)	44 (15%)	47 (16%)
>=80	62 (20%)	75 (25%)	56 (18%)	53 (17%)	48 (16%)
Diagnosis	. ,		, , , , , , , , , , , , , , , , , , ,	, ,	, , , , , , , , , , , , , , , , , , ,
Mechanical Complication	172 (57%)	198 (65%)	208 (69%)	212 (70%)	192 (63%)
Fracture	26 (9%)	29 (10%)	18 (6%)	10 (3%)	31 (10%)
Progressive OA	39 (13%)	32 (11%)	24 (8%)	24 (8%)	18 (6%)
Hospital Frailty Risk Score					
None	158 (52%)	115 (38%)	149 (49%)	156 (51%)	158 (52%)
Mild	100 (33%)	123 (41%)	106 (35%)	100 (33%)	99 (33%)
Moderate	38 (13%)	56 (18%)	40 (13%)	43 (14%)	36 (12%)
Severe	8 (3%)	9 (3%)	8 (3%)	4 (1%)	10 (3%)
Annual Surgeon Volume					
Volume 0-4	30 (10%)	39 (13%)	31 (10%)	30 (10%)	24 (8%)
Volume 5-9	43 (14%)	46 (15%)	44 (15%)	41 (14%)	35 (12%)
Volume 10-14	89 (29%)	72 (24%)	55 (18%)	63 (21%)	58 (19%)
Volume 15-19	49 (16%)	56 (18%)	44 (15%)	47 (16%)	53 (17%)
Volume 20-24	48 (16%)	64 (21%)	74 (24%)	62 (20%)	47 (16%)
√olume >=25	45 (15%)	26 (9%)	55 (18%)	60 (20%)	86 (28%)
Perioperative Outcomes					
Readmission within 30 days	15(5%)	36(12%)	23(8%)	17(6%)	20(7%)
90 Day Mortality	8(3%)	15(5%)	10(3%)	7(2%)	10(3%)
Prolonged Length of Stay	135(44%)	134(44%)	135(45%)	130(43%)	132(44%)

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Table 2 – Multivariable Logistic Regression showing Odds Ratios for primary and secondary outcomes by straight line travel quintile

	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Readmission with 30	2.23 (95% CI 1.19 to	1.55 (95% CI 0.79 to	1.06 (95% CI 0.51 to	1.44 (95% CI 0.71 to
days	4.37),p=0.01	3.13),p=0.21	2.23),p=0.87	2.96),p=0.32
90 Day Mortality	1.79 (95% CI 0.66 to	1.55 (95% Cl 0.52 to	1.72 (95% CI 0.53 to	1.46 (95% CI 0.49 to
	5.17),p=0.27	4.70),p=0.43	5.64),p=0.36	4.53),p=0.50
Prolonged Length of	0.90 (95% CI 0.62 to	1.02 (95% Cl 0.71 to	0.99 (95% CI 0.69 to	0.96 (95% CI 0.67 to
stay	1.30),p=0.57	1.46),p=0.91	1.412),p=0.95	1.39),p=0.84

Table 3 – Sensitivity analysis exploring road travel distance quintile and primary/secondary outcomes

	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Readmission with 30 days	2.11 (95% CI 1.14 to 4.06),p=0.02	1.41 (95% CI 0.71 to 2.84),p=0.33	1.44 (95% CI 0.73 to 2.90),p=0.29	1.16 (95% CI 0.56 to 2.4),p=0.68
90 Day Mortality	1.52 (05% CL0.58 to	1 18 (95% CI 0 39 to	2 42 (95% CL 0 83 to	0.83 (95% CI 0.25 to
Jo Day Mortanty	4.21),p=0.40	3.56),p=0.77	7.27),p=0.11	2.64),p=0.75
Prolonged Length of stay	0.69 (95% CI 0.48 to 0.99),p=0.04	0.94 (95% CI 0.66 to 1.35),p=0.74	0.86 (95% CI 0.60 to 1.23),p=0.41	0.95 (95% CI 0.66 to 1.36),p=0.77

Table 4- Sensitivity analysis exploring driving time quintile and primary/secondary outcomes

	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Readmission with 0 days	1.19 (95% CI 0.63 to 2.25),p=0.59	1.23 (95% CI 0.65 to 2.35),p=0.52	1.53 (95% CI 0.82 to 2.89),p=0.18	0.92 (95% CI 0.45 to 1.85),p=0.81
0 Day Mortality	2.29 (95% CI 0.84 to 6.76),p=0.12	1.52 (95% CI 0.50 to 4.74),p=0.46	2.63 (95% CI 0.90 to 8.16),p=0.08	0.94 (95% CI 0.26 to 3.23),p=0.91
rolonged Length f stay	0.82 (95% CI 0.57 to 1.52),p=0.27	0.80 (95% CI 0.56 to 1.15),p=0.23	1.27 (95% CI 0.89 to 1.81),p=0.19	1.01 (95% CI 0.70 to 1.44),p=0.97
		(2	

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

3	
4	
5	
6	
7	
, Q	
0	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
10	
עו רכ	
20	
21	
22	
23	
24	
25	
26	
27	
28	
20	
29	
20	
31	
32	
33	
34	
35	
36	
37	
38	
30	
10	
40 1	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
51	
52 52	
22	
54	
55	
56	
57	
58	
59	

1 2

Supplementary Material

Code	Code description
OPCS-4	codes for knee revision procedures
O180	Conversion from previous hybrid prosthetic replacement of knee joint using cement
O182	Conversion to hybrid prosthetic replacement of knee joint using cement
O183	Revision of hybrid prosthetic replacement of knee joint using cement
O184	Attention to hybrid prosthetic replacement of knee joint using cement
W400	Conversion from previous cemented total prosthetic replacement of knee joint
W402	Conversion to total prosthetic replacement of knee joint using cement
W403	Revision of total prosthetic replacement of knee joint using cement
W404	Revision of one component of total prosthetic replacement of knee joint using cement
W410	Conversion from previous uncemented total prosthetic replacement of knee joint
W412	Conversion to total prosthetic replacement of knee joint not using cement
W413	Revision of total prosthetic replacement of knee joint not using cemen
W414	Revision of one component of total prosthetic replacement of knee joint not using cement
W420	Conversion from previous total prosthetic replacement of knee joint NEC

3 4	W422	Conversion to total prosthetic replacement of knee joint NEC
5 6	W423	Revision of total prosthetic replacement of knee joint NEC
7 8 9	W424*	Attention to total prosthetic replacement of knee joint NEC
10 11 12 13	W425	Revision of one component of total prosthetic replacement of knee joint NEC
14 15 16 17	W522†	Conversion to prosthetic replacement of articulation of bone using cement NEC
18 19 20 21	W523†	Revision of prosthetic replacement of articulation of bone using cement NEC
22 23 24 25	W532†	Conversion to prosthetic replacement of articulation of bone not using cement NEC
26 27 28 29	W533†	Revision of prosthetic replacement of articulation of bone not using cement NEC
30 31 32	W542†	Conversion to prosthetic replacement of articulation of bone NEC
33 34	W543†	Revision of prosthetic replacement of articulation of bone NEC
35 36 37	W544*†	Attention to prosthetic replacement of articulation of bone NEC
38 39	W553†	Conversion to prosthetic interposition arthroplasty of joint
40 41	W564†	Conversion to interposition arthroplasty of joint NEC
42 43 44	W574†	Conversion to excision arthroplasty of joint
45 46	W582†	Revision of resurfacing arthroplasty of joint
47 48 49	W603†	Conversion to arthrodesis and extra-articular bone graft NEC
50 51	W613†	Conversion to arthrodesis and articular bone graft NEC
52 53	W641†	Conversion to arthrodesis and internal fixation NEC
55 56	W642†	Conversion to arthrodesis and external fixation NEC
57 58 59	OPCS-4 c	odes for laterality

Z941	Bilateral
Z942	Left-sided
Z943	Right-sided
ICD-10 co	odes for Infection
T845	Infection and inflammatory reaction due to internal joint prosthesis
T846	Infection and inflammatory reaction due to internal fixation device [any site]
T847	Infection and inflammatory reaction due to other internal orthopaedic prosthetic devices, implants and grafts
T814	Infection following a procedure, not elsewhere classified
ICD-10 co	odes for fracture
M966	Fracture of bone following insertion of orthopaedic implant, joint
	prosthesis or bone plate
ICD-10 co	odes for mechanical complications
T840	Mechanical complication of internal joint prosthesis
T841	Mechanical complication of internal fixation device of bones of limb
T842	Mechanical complication of internal fixation device of other bones
T843	Mechanical complication of other bone devices, implants and grafts
T844	Mechanical complication of other internal orthopaedic devices, imnplants and grafts
ICD-10 co	odes for osteoarthritis/arthrosis
	Polyarthrosis
M15-	
M15- M17-	Gonarthrosis

Classification of Diseases and Related Health Problems, tenth revision. * Where OPCS-4 codes Y032 (renewal of prosthesis in organ NOC) or Y037 (removal of prosthesis from organ NOC) were also used. † Where OPCS-4 codes O132 (knee NEC) or Z765 (lower end of femur NEC) or Z774 (upper end of tibia NEC) or Z787 (patella) or Z844 (patellofemoral joint) or Z845 (tibiofemoral joint) or Z846 (knee joint) or Z851 (upper tibiofibular joint) were used to identify knee as the body site.

Supplementary material S2 – Flow of patient inclusion/exclusions



Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Supplementary material S3 - R code

See separate .R file

to beet terien only
BMJ Open

What is the impact of longer patient travel distances and times on perioperative outcomes following revision knee replacement for aseptic reasons: An analysis using national administrative data from Hospital Episode Statistics

Journal:	BMJ Open
Manuscript ID	bmjopen-2024-085201.R1
Article Type:	Original research
Date Submitted by the Author:	16-Oct-2024
Complete List of Authors:	Matthews, Alexander; University of Exeter Faculty of Health and Life Sciences, Department of Clinical and Biomedical Sciences; Royal Devon and Exeter NHS Foundation Trust, Evans, Jonathan; University of Exeter Medical School, Health Services and Policy Research; Royal Devon and Exeter NHS Foundation Trust, Princess Elizabeth Orthopaedics Centre Evans, Jonathan; University of Exeter, Department of Public Health and Sports Sciences Lamb, Sarah; University of Exeter, College of Medicine and Health Price, Andrew; University of Oxford, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences Gray, William; NHS England and NHS Improvement London Briggs, Tim; NHS England and NHS Improvement London Toms, Andrew ; Princess Elizabeth Orthopaedic Centre
Primary Subject Heading :	Public health
Secondary Subject Heading:	Surgery
Keywords:	Knee < ORTHOPAEDIC & TRAUMA SURGERY, PUBLIC HEALTH, Health Services

SCHOLARONE[™] Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our <u>licence</u>.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which <u>Creative Commons</u> licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

terez oni

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

Title

What is the impact of longer patient travel distances and times on perioperative outcomes following revision knee replacement for aseptic reasons: An analysis using national administrative data from Hospital Episode Statistics

Names, Affiliations, and positions of all authors

Alex Matthews1,2,3,4, Jonathan P Evans2,3, Jonathan T Evans2,3, Sarah E Lamb3 Andrew Price4,5, William Gray1, Tim Briggs1,6, Andrew Toms2,3

- 1. Getting It Right First Time programme, NHS England, London, UK
- 2. Royal Devon University Healthcare NHS Foundation Trust, Exeter, UK
- 3. University of Exeter, Exeter, UK
- 4. Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, UK

5. Nuffield Orthopaedic Centre, Oxford, UK

6. Royal National Orthopaedic Hospital, Stanmore, London, UK

Correspondence: Alexander Matthews, Getting It Right First Time programme, NHS England and NHS Improvement, Wellington House, Waterloo Road. London, UK, SE1 8UG; telephone: 07401932627 email: alexander.matthews1@nhs.net

Word count: [3316 words]

Short title: Patient Journey Distance and Time for Complex Revision Knee Surgery

Key words: social determinants of health, revision knee replacement, contextual factors in complex health interventions

Twitter handles:

Alex Matthews: @DrAlexMatthews

Tim WR Briggs: @ProfTimBriggs

WilliamKGray1: @WilliamKGray1

GIRFT programme: @NHSGIRFT

Structured Abstract (Word count suggested 250-300)

Objectives

Patients with problematic knee replacements requiring further surgery often have difficulties mobilising and increasingly rely on family support. Evolving practice in England aims to manage these patients in specialised centres with the intention of improving outcomes. This practice will result in longer travel distances and times in this frailer group of patients. We want to examine the types of distances and travel times patients can be expected to travel for complex orthopaedic surgery and to explore concerns of how these impact patient outcomes.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

Design

Retrospective observational study from the Hospital Episode Statistics. Multivariable adjusted logistic regression modelling was used to compare the exposure variable with perioperative outcomes

Setting

Patients presenting to tertiary referral centres between 1st January 2016 to 31st December 2019. A tertiary referral centre was defined as a trust performing >70 revisions in the year prior.

Participants

Adult patients undergoing revision total knee replacement procedures for aseptic reasons between 1st January 2016 to 31st December 2019.

Interventions

Patient level travel distance and time was calculated using the department of health Journey Time Statistics.

Main Outcome Measures

The primary outcome is the association of travel distance and time on emergency readmission within 30 days. Secondary outcomes will focus on mortality within 90 days and length of inpatient stay.

Results

1516 patients were treated at 16 tertiary referral centres for non-infected reasons. Patients in the longest driving distance group were expected to travel a median distance of 44.55 miles (IQR 35.90 to 56.30) with an expected median journey time of 66.3 minutes (IQR 57.9 to 80.5). Overall, 30-day readmission was not statistically associated with farther travel distances or driving times.

Conclusions

Patients were expected to travel up to hour for revision knee replacement surgery. There was no association between increasing travel distance and time on perioperative outcomes.

Strengths and limitations of this study

- Our study is the first to describe travel distance and time associations using a large revision knee replacement sample providing data across multiple years
- This data reflects revision knee replacement procedures undertaken across different geographical areas of England
- Owing to differences in the coverage of Hospital Episode Statistics, procedures in hospitals outside of England were not included in this analysis
- Clinical coding practice within HES is known to vary between trusts but this is unlikely to be vary systematically to bias our findings
- This analysis only reports travel times for patients with access to their own transport and does not consider times for those patients using public transport

Introduction

Primary knee replacement is a successful procedure that improves quality of life for the majority of patients.[1] However, at 10 years following a primary knee replacement, about 3.5% of patients will have undergone a revision surgery.[2] The majority of these procedures are carried out due to infection or polyethylene wear of the implant.[3] A failed primary knee replacement represents a life changing transition point where individuals are likely to suffer from pain, reduced mobility as well as dependency on family members.[4] Patients often face multi-step surgery with longer hospital length of stays and higher complication rates.[5, 6]

The orthopaedic GIRFT (Getting It Right First Time) programme was launched in 2012 following the publication of the Orthopaedic National Report.[7] A key recommendation was the centralisation of complex orthopaedic surgery, including revision knee surgery, to specialist centres with the aim of improved patient outcomes. Consequently, revision total knee replacement (rTKR) surgery in the England has evolved into a regional network service model.[8] In doing so, all hospitals performing rTKRs form a network in the respective regions. Less specialist hospitals defined by lower annual case volume thresholds are encouraged to discuss and sometimes refer their caseload to more specialist centres. Several studies based on large revision hip and knee registries has suggested this model carries a lower failure rate defined by the need for further revision surgery.[9-11] Early evidence has suggested reduced early failure rates through the adoption of revision knee networks.[12]

However, this approach to managing patients is inevitably associated with increasing travel distances between some patient's homes and their treating hospital. Expected distances are important to explore, particularly as these patients struggle with mobility, may be unable to drive and may be more reliant on family members.[4] Furthermore, greater travel distances have been associated with higher readmission rates and higher mortality rates following complex vascular surgery.[13] The pick-up rate of early complications, avoiding the need for readmission, may be less in areas further away from the main treatment centre. There is also concern that patients

 required to travel greater distances are more likely to be re-admitted to a different hospital resulting in clinical decisions that do not incorporate the primary surgeon and potentially alter outcomes.[14] Subsequently the aims of this paper is to examine if the same association with longer patient travel distance and perioperative outcomes exists following complex orthopaedic surgery with a focus on revision knee replacement surgery performed in high volume tertiary referral centres.

Methods

Design

This study is a retrospective data analysis of observational data from the Hospital Episode Statistics (HES) and Office for National Statistics (ONS) databases. HES data is collected by NHS England for all patients treated at NHS hospitals in England and those treated at private hospitals where treatment was funded by the NHS. This study complies with the recommended reporting guidelines when using HES data[15] and the Strengthening of Reporting of Observational studies in Epidemiology (STROBE) guidelines.[16]

The analysis and presentation of data follows current NHS England guidance for the use of HES data for research purposes[17] and is anonymised to the level required by ISB1523 Anonymisation Standard for Publishing Health and Social Care Data.[18] The HES data were linked at a patient level to data from the ONS on deaths and date of death, which allowed the identification of patients who had died after their surgery. Linkage was achieved using a unique patient identifier using a previously validated methodology.^[19]

Patient travel distances were calculated using the Journey Time Statistics reference document produced by the UK Department of Transport which modelled theoretical journey times between known Lower Layer Super Output Areas (LSOA) of residence and NHS hospital sites.[20] The Journey Time Statistics document is available in the supplementary material section.

Population

An rTKR procedure was defined as a permanent removal or exchange of knee arthroplasty components. This includes a revision of a total knee replacement and a conversion of a unicondylar knee replacement to a total knee replacement. Secondary patellar resurfacing was not included as this represents a simple revision procedure, one that can be carried out in most nonspecialised hospitals. All patients aged \geq 18 years who underwent a rTKR in a high-volume trust between 1st January 2016 to 31st December 2019 were included in the study population. A high-volume trust was classified as a centre performing > 70 revisions per year. This revision volume threshold for classification represents those proposed by the British Association for Surgeons of the Knee (BASK) Revision Knee Working Group. [21] These are more likely to represent tertiary referral centres where the stratification of more complex work will take place. Annual case volume at each trust was defined as the number of revision cases conducted in the year prior to the index procedure. This measure was preferred over a simple calculation of average annual volume as it accounts for recent experience at the point of surgery. The Office for Population Censuses and Surveys' Classification of Interventions and Procedures version 4 (OPCS-4) codes used to identify rTKR procedures are detailed in **Supplementary** material S1. Where the procedure laterality was not specified, patients were excluded. The flow of patients, with numbers excluded at each point, is summarised in **Supplementary material S2**. To manage population heterogeneity, data were extracted for the period 1st April 2011 to 31st December 2019 and only the first revision for a specific side of the body record in this time period included. [22] Thus, any early revisions on the same side of the body in the four years and nine months preceding the start of the study period were identified and these patients excluded from the study. This aims to exclude the early revision knee replacement failures which have been shown to represent catastrophic failures potentially skewing our results. [22]We excluded revisions for infection as these represent a more variable patient group with a different complication profile [23] and this is further discussed in our study limitations.

Exposure variable

In the analysis straight line travel distance was calculated between a patient's LSOA and the postal codes for the treating hospitals. LSOAs are determined by the Office for National Statistics and are designed for the reporting of small area statistics. Public transport and highways data for England was used to create theoretical journey distances and times from origins to destinations. The resulting travel distances and/or times for each patient were divided into quintiles a priori, following a recently reported methodology.[13] Sensitivity analyses were performed using travel distances by road and peak driving times to account for variation in travel infrastructure between rural and urban areas and to attribute more meaningful results for patients. Peak driving times were calculated by using average traffic speeds for between 7am and 10am.

Outcomes

The primary outcome was emergency readmission within 30 days of discharge from the index surgical hospital. Secondary outcomes included 90-day mortality, and hospital length of stay (LOS) above the median. The LOS outcome was dichotomised into above median or below median LOS of five days.

Statistical Analyses

Data was extracted from a secure, encrypted server controlled by NHS Digital. Data were analysed within a secure, encrypted environment using standard statistical software: R Studio version 2023.09.1+494 (Boston, Massachusetts, USA). The R code and packages used are included in **Supplementary material S3**

Crude comparisons of baseline categorical characteristics and travel distance proximity were calculated. A This data were categorical in nature and summarised as frequency and percentage. In primary analysis a logistic multivariable regression model was constructed to evaluate associations between travel distance quintiles and 30 day readmission, with adjustment for the covariates listed above. The first (shortest) travel distance quintile was used as the reference in all models.

Age, sex , comorbidities and characteristics of initial presentation were included in the logistic regression model. These variables have been shown to influence the risk of complications after R-TKA and therefore represent known confounders.[9, 10, 23]

This also included data on economic deprivation measured using the Index of Multiple Deprivation (IMD).[24] The IMD gives the LSOA where the patient lives a score based on a range of measures of deprivation. IMD was categorised into quintiles, based on all-England data, for analysis. A spearman's rank correlation was performed to investigate the relationship between IMD score and travel distances. Health co-morbidity was quantified using the Hospital Frailty Risk Score (HFRS). HFRS identifies frailty based on the occurrence of any of 109 International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10) codes used during any hospital admissions in the two years prior to, and for, the index admission.

All secondary outcomes were binary and analysed using the same multivariable logistic regression. Multicollinearity was assessed with reference to variance inflation factor and Shapiro-Wilcox test of normality. Model fit was assessed with reference to the pseudo R² values.

A supplementary analysis is available analysing travel times and distances as a continuous variable with the primary outcome. Please see supplementary material S4

Results

Demographic characteristics and co-morbidities

The 1,516 patients in the overall study population, were divided into quintiles of travel distance of 303 or 304 patients each. The median straight line travel distance for quintile one was 2.3 miles (IQR 1.3 to 3.1). For the fifth travel quintile, median distance was 33.5 miles (IQR 25.5 to 41.1). Baseline co-morbidities and demographic characteristics were broadly similar among the travel distance quintiles (**Table 1**). Travel distance was not strongly correlated with age or social deprivation (**Figure 1**)

Table 1 – Baseline demographics and clinical characteristics and rawperioperative outcomes for patients by travel distances quintile

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Number of Patients	304	303	303	303	303

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
50
59
00

Deprivation Quintile					
1(Most Deprived)	54 (18%)	87 (29%)	50 (17%)	58 (19%)	57 (19%
2	48 (16%)	61 (20%)	68 (22%)	58 (19%)	66 (22%)
3	57 (19%)	65 (21%)	63 (21%)	63 (21%)	56 (18%)
4	55 (18%)	43 (14%)	49 (16%)	80 (26%)	75 (25%
5 (Least Deprived)	90 (30%)	47 (16%)	73 (24%)	44 (15%)	49 (16%
Sex					
Male	121 (40%)	133 (44%)	135 (45%)	124 (41%)	141 (47%
Age in years					
16-59	52 (17%)	42 (14%)	56 (18%)	57 (19%)	68 (22%
60-64	46 (15%)	29 (10%)	35 (12%)	38 (13%)	47 (16%
65-69	53 (17%)	58 (19%)	56 (18%)	52 (17%)	44 (15%
70-74	45 (15%)	56 (18%)	53 (17%)	59 (19%)	49 (16%
75-79	46 (15%)	43 (14%)	47 (16%)	44 (15%)	47 (16%
>=80	62 (20%)	75 (25%)	56 (18%)	53 (17%)	48 (16%
Diagnosis					
Mechanical Complication	172 (57%)	198 (65%)	208 (69%)	212 (70%)	192 (63%
Fracture	26 (9%)	29 (10%)	18 (6%)	10 (3%)	31 (10%
Progressive OA	39 (13%)	32 (11%)	24 (8%)	24 (8%)	18 (6%)
Hospital Frailty Risk Score					
None	158 (52%)	115 (38%)	149 (49%)	156 (51%)	158 (52%
Mild	100 (33%)	123 (41%)	106 (35%)	100 (33%)	99 (33%
Moderate	38 (13%)	56 (18%)	40 (13%)	43 (14%)	36 (12%
Severe	8 (3%)	9 (3%)	8 (3%)	4 (1%)	10 (3%)
Annual Surgeon Volume					
Volume 0-4	30 (10%)	39 (13%)	31 (10%)	30 (10%)	24 (8%)
Volume 5-9	43 (14%)	46 (15%)	44 (15%)	41 (14%)	35 (12%)
Volume 10-14	89 (29%)	72 (24%)	55 (18%)	63 (21%)	58 (19%
Volume 15-19	49 (16%)	56 (18%)	44 (15%)	47 (16%)	53 (17%
Volume 20-24	48 (16%)	64 (21%)	74 (24%)	62 (20%)	47 (16%
Volume >=25	45 (15%)	26 (9%)	55 (18%)	60 (20%)	86 (28%
Perioperative Outcomes		-			
Readmission within 30 days	15(5%)	36(12%)	23(8%)	17(6%)	20(7%)
90 Day Mortality	8(3%)	15(5%)	10(3%)	7(2%)	10(3%)
	135(11%)	134(44%)	135(45%)	130(43%)	132/1/19

Association between travel distance and readmission, mortality and extended hospital stay

Overall, 111 patients were readmitted within 30 days. Crude comparisons of proportions readmitted within 30 days for each travel distance quintiles revealed a higher rate of readmission for the second travel quintile. In multivariable adjusted logistic regression, there was no statistically significant association between travel distance and readmission within 30 days (**Table 2**). Odds for 30-day readmission was 1.44 (95% CI 0.71 to 2.96, P 0.32) for Q5 compared with Q1. Increased travel distance was not associated with a significant change in the odds of death within 90 days (OR for Q5 vs Q1, 1.46 (95% CI 0.49 to 4.53, P 0.682)). Travel distance quintile was not associated with prolonged length of hospital stay related to the index surgery after multivariable adjustment (OR for Q5 vs Q1, 0.96 (95% CI 0.67 to 1.39, P 0.84)).

Table 2 – Multivariable Logistic Regression showing Odds Ratios for primary and secondary outcomes by straight line travel quintile

	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Readmission with 30	2.23 (95% CI 1.19 to	1.55 (95% CI 0.79 to	1.06 (95% CI 0.51 to	1.44 (95% CI 0.71 to
days	4.37),p=0.01	3.13),p=0.21	2.23),p=0.87	2.96),p=0.32
90 Day Mortality	1.79 (95% CI 0.66 to 5.17),p=0.27	1.55 (95% CI 0.52 to 4.70),p=0.43	1.72 (95% CI 0.53 to 5.64),p=0.36	1.46 (95% CI 0.49 to 4.53),p=0.50
Prolonged Length of stay	0.90 (95% CI 0.62 to 1.30),p=0.57	1.02 (95% CI 0.71 to 1.46),p=0.91	0.99 (95% CI 0.69 to 1.412),p=0.95	0.96 (95% CI 0.67 to 1.39),p=0.84

Real world travel distance and outcomes

The above results used straight line travel distance between patient' LSOA and treating hospital. A sensitivity analysis using actual patient travel distances using the shortest possible road route was performed **(Table 3)**. The median driving distance by the shortest possible road route for the closest quintile was 3.40 miles (IQR 2.00 to 4.40). The furthest quintile median driving distance was 44.55 miles (IQR 35.90 to 56.30) This analysis showed no association between driving distance and all perioperative outcomes. The OR for readmission within 30 days in Q5 vs Q1 was 1.16(0.56-2.41, p value = 0.68).

Table 3 – Sensitivity analysis exploring road travel distance quintile and primary/secondary outcomes

Quintile 3

Quintile 2

Quintile 4

Quintile 5

Readmission with	2.11 (95% CI 1.14 to	1.41 (95% CI 0.71 to	1.44 (95% CI 0.73 to	1.16 (95% CI 0.56 to
30 days	4.06),p=0.02	2.84),p=0.33	2.90),p=0.29	2.4),p=0.68
90 Day Mortality	1.52 (95% CI 0.58 to 4.21),p=0.40	1.18 (95% CI 0.39 to 3.56),p=0.77	2.42 (95% CI 0.83 to 7.27),p=0.11	0.83 (95% CI 0.25 to 2.64),p=0.75
Prolonged Length	0.69 (95% CI 0.48 to 0.99),p=0.04	0.94 (95% CI 0.66 to	0.86 (95% CI 0.60 to	0.95 (95% CI 0.66 to
of stay		1.35),p=0.74	1.23),p=0.41	1.36),p=0.77

Journey drive times and outcomes

A further sensitivity analysis using driving times was calculated (Table 4). The median drive time for quintile one was 12.6 minutes (IQR 8.7 to 15.3). For the fifth time quintile the median was 66.3 minutes (IQR 57.9 to 80.5). No statistical association was found between drive time and perioperative outcomes. The OR for readmission within 30 days in Q5 vs Q1 was 0.92 (0.45 - 1.85, p value = 0.81)

Table 4- Sensitivity analysis exploring driving time quintile and

primary/secondary outcomes

	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Readmission with 30 days	1.19 (95% CI 0.63 to 2.25),p=0.59	1.23 (95% CI 0.65 to 2.35),p=0.52	1.53 (95% CI 0.82 to 2.89),p=0.18	0.92 (95% CI 0.45 to 1.85),p=0.81
90 Day Mortality	2.29 (95% CI 0.84 to 6.76),p=0.12	1.52 (95% CI 0.50 to 4.74),p=0.46	2.63 (95% CI 0.90 to 8.16),p=0.08	0.94 (95% CI 0.26 to 3.23),p=0.91
Prolonged Length of stay	0.82 (95% CI 0.57 to 1.52),p=0.27	0.80 (95% CI 0.56 to 1.15),p=0.23	1.27 (95% CI 0.89 to 1.81),p=0.19	1.01 (95% CI 0.70 to 1.44),p=0.97

Discussion

Statement of principal findings

We present a multi-hospital site retrospective analysis of patients undergoing revision knee replacement surgery at tertiary referral centres in England. In this analysis of 1,516 patients undergoing aseptic revision knee replacement surgery, we did not observe an association between distance and time travelled for revision surgery and readmission within 30 days. Patients in the longest driving time category were expected to travel for a median time of more than one hour.

Strengths and weaknesses of the study

The findings of this study should be interpreted in view of several limitations. Firstly, this analysis used observational data from a large administrative dataset covering all NHS-funded procedures conducted in England. As with all datasets we are limited in the amount of detail provided regarding presentation. We chose to categorise a highvolume centre by trust to accurately capture surgical experience. All NHS hospitals in England are run by hospital trusts which typically involve between one and four hospitals within a catchment area standardising their practice. It is common practice for specialist orthopaedic surgeons to move between these sites delivering the same procedures. Our study involved 35 hospital sites run by 16 trusts. We acknowledge this is a weakness of our study as this may not be representative of all trusts. The indication for revision coded as mechanical complication encompasses several common indications such as aseptic loosening, instability and malalignment. Reassuringly these indications have similar length of stay, and perioperative outcomes.[23] Differences exist in their re-revision rate, however this was not an outcome of focus in our study. It is likely that the complexity of the surgery undertaken may vary within the different indications for revision. Evidence suggests that operative surgical time is related to increased length of stay in aseptic revision knee replacement.[25]. There is a lack of granular data for revisions due to infection and therefore we excluded this patient group as some readmissions for this patient group may represent planned readmissions. There is also a lack of granular clinical data using HES for each readmission, therefore we cannot ascertain precise reasons

for readmissions, but we assume are related to a post-surgical complication. Clinical coding practice within HES is known to vary across trusts.[26] As an example, some trusts may be more consistent in coding comorbidities, and this may have created some bias. However, this is unlikely to vary systematically with travel distances and so significantly bias our findings. We acknowledge the relatively short travel distances in this population compared to examples from the United States as such the results of this study may not be generalisable to larger geographical areas or less mature healthcare systems. However, the upper quintile in our study represents a substantial journey distance and time for our patient cohort where poor mobility is a significant factor affecting their care. This analysis does not consider journey times of those who may not have access to a car and instead chose to take public transport.

Strengths and weaknesses in relation to other studies, discussing important differences in results

This is the first study to analyse the potential impact of patient travel distances on patients receiving complex orthopaedic surgery. The findings that longer travel distances are not associated with inferior outcomes is an important part of the evaluation of the assumptions and context behind the establishment of revision knee networks.[27] This study has shown that concerns of introducing a network in larger geographical regions, for example in Scotland where longer patient travel distances and times are common, may be less important.[28] This is particularly useful as regions explore the geography of their revision networks and during summative outcome assessment of this complex health intervention.[29]

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

It may be seen as surprising that no association between travel distance and prolonged length of hospital stay was identified. An expectation exists of increasing difficulties being encountered with the discharge of patients living greater distances from their treating hospital, which has been observed in patients following elective pancreatic surgery.[30] This is also an observation seen in patients being treated in specialist vascular centres in the United States which led to the recommendation of additional care coordination and follow up efforts. However, the geography of the population in these studies was much larger with significantly longer travel distances.

We did not observe a strong correlation between social deprivation status and age of the patient with longer travel distances. It is reassuring that access to treatment for older patients and those from poor socioeconomic backgrounds is unaffected by travel distance. However, there may be patients who refused to travel to a specialist centre and opted for treatment at their local centre.

Meaning of the study: possible explanations and implications for clinicians and policymakers

The organisation and delivery of revision knee services in England has recently undergone a substantial change and now such services are provided around regional networks of care. This promises substantial advantages to the increasing number of patients with problematic knee replacements in our ageing population who will benefit from regional expertise.[8] However, it is unknown the impact of patients residing farther from tertiary referral centres, particularly rural patients who may encounter additional difficulties associated with greater travel distance. A recent study following the outcomes of aortic surgery found that longer travel distances are associated with inferior perioperative outcomes[13]. Similar associations have been found in postoperative colorectal surgery patients [31]. As such our results are reassuring to policy makers and clinicians.

Unanswered questions and future research

There is a scarcity of evidence evaluating the patient perception of complex health interventions such as network models of care. Recent work by Kugler et al has demonstrated the willingness of patients to travel further for better outcomes in the context of total knee replacement surgery. [32] Nevertheless, patient perceptions of travelling further for their treatment should be a focus for future research in the context of revision knee patients, particularly as this is one of the top ten research priorities identified by the James Lind Alliance priority setting partnership.[33]

Conclusion

We did not observe an association in our study population between 30-day readmission rates and increasing travel distances or times between a patient's home and their treating hospital in revision knee replacement. This paper is the first to explore the relationship between travel distance and complex orthopaedic surgery and informs some concerns regarding the creation of a centralised revision knee network. This information is of utility to surgical providers and commissioners of healthcare services. Furthermore, it can inform patient-led decision making and the exploration of perceptions surrounding travelling for complex surgery. Although this is the first assessment in complex orthopaedic surgery, a prospective analysis will be undertaken as part of the ongoing auditing of revision knee networks in England.

Supplementary material and figures

Supplementary material S1 – OPCS-4 code criteria used for Hospital Episode Statistics data extraction

See separate file named supplementary material S1

Supplementary material S2 – Flow of patient inclusion/exclusions

-See attached file named Supplementary Material S2

Supplementary material S3 - R Code

See attached file named Supplementary Material S3

Supplementary material S4 – Relationship between Travel distances and times modelled as a continuous variable with primary outcome (readmission within 30 days)

See attached file named supplementary Material S4

Figure 1 – Box plot showing association of social deprivation and age on travel distance quintile. Spearman's rank correlation investigating relationship between these factors with travel distance.

See attached files called 'Figure 1 – Deprivation and Travel Distance' AND 'Figure 1 – Age and Travel Distance'

Contributorship

Alex Matthews: Conceptualisation, Methodology, Project Administration, Investigation, Data Curation, Formal Analysis, Visualisation, Writing - original draft, Writing - review and editing. This author is the guarantor and is responsible for the content

Jonathan P Evans: Conceptualisation, Supervision, Writing - review & editing

Jonathan T Evans: Supervision, Writing - review and editing

Sarah E Lamb: Conceptualisation, Supervision, Writing - review and editing

Andrew Price: Conceptualisation, Supervision, Writing - review and editing

William Gray: Conceptualisation, Supervision, Methodology, Writing - review and editing

Tim Briggs: Supervision, Writing - review and editing

Andrew Toms: Conceptualisation, Supervision, Writing - review and editing

Public and Patient Involvement statement

The study's chief investigator (AT) led the James Lind Alliance 'Revision Knee Replacement' priority setting partnership. This group of patients, carers and health care professionals identified the need to investigate the best way of organising revision knee replacement surgery to improve patient outcomes as one of their top 10 research questions. Patients were therefore directly involved in the development of the study's aims and objectives. The results of the study will be disseminated to the members of this group prior to publication.

Copyright/licence for publication Statement

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence.

Competing Interests

All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

Transparency Declaration

 The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Ethical Approval

The analysis and presentation of data follows current NHS England guidance for the use of HES data for research purposes and is anonymised to the level required by ISB1523 Anonymisation Standard for Publishing Health and Social Care Data. Ethical approval was not required.

Funding

No funding was obtained to carry out this study

Data Sharing

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

References

1. Price AJ, Alvand A, Troelsen A, Katz JN, Hooper G, Gray A, et al. Knee replacement. Lancet. 2018;392(10158):1672-82.

2. Sabah SA, Hedge EA, Abram SGF, Alvand A, Price AJ, Hopewell S. Patient-reported outcome measures following revision knee replacement: a review of PROM instrument utilisation and measurement properties using the COSMIN checklist. BMJ Open. 2021;11(10):e046169.

3. Sabah S, Knight R, Alvand A, Murray D, Petrou S, Beard D, Price A. No exponential rise in revision knee replacement surgery over the past 15 years: an analysis from the National Joint Registry. Osteoarthritis and Cartilage. 2022;30(12):1670-9.

4. Rowland T, Lindberg-Larsen M, Santy-Tomlinson J, Jensen CM. A qualitative study of patients' experiences before, during and after surgical treatment for periprosthetic knee joint infection; "I assumed it had to be like that ... ". International Journal of Orthopaedic and Trauma Nursing. 2023;48:100992.

5. Garbarino LJ, Gold PA, Sodhi N, Anis HK, Ehiorobo JO, Boraiah S, et al. The effect of operative time on in-hospital length of stay in revision total knee arthroplasty. Ann Transl Med. 2019;7(4):66.

6. Liodakis E, Bergeron SG, Zukor DJ, Huk OL, Epure LM, Antoniou J. Perioperative Complications and Length of Stay After Revision Total Hip and Knee Arthroplasties: An Analysis of the NSQIP Database. The Journal of Arthroplasty. 2015;30(11):1868-71.

7. Briggs TWR. A National Review of Adult Elective Orthopaedic Services in England. In: Time GIRF, editor. 2015.

8. Kalson NS, Mathews JA, Phillips JRA, Baker PN, Price AJ, Toms AD, et al. Revision knee replacement surgery in the NHS: A BASK surgical practice guideline. The Knee. 2021;29:353-64.

9. Yapp LZ, Walmsley PJ, Moran M, Clarke JV, Simpson AHRW, Scott CEH. The effect of hospital case volume on re-revision following revision total knee arthroplasty. The Bone & Joint Journal. 2021;103-B(4):602-9.

10. Halder AM, Gehrke T, Günster C, Heller KD, Leicht H, Malzahn J, et al. Low Hospital Volume Increases Re-Revision Rate Following Aseptic Revision Total Knee Arthroplasty: An Analysis of 23,644 Cases. J Arthroplasty. 2020;35(4):1054-9.

11. Jeschke E, Gehrke T, Günster C, Heller KD, Leicht H, Malzahn J, et al. Low Hospital Volume Increases Revision Rate and Mortality Following Revision Total Hip Arthroplasty: An Analysis of 17,773 Cases. J Arthroplasty. 2019;34(9):2045-50.

12. Bloch BV, Matar HE, Berber R, Gray WK, Briggs TWR, James PJ, Manktelow ARJ. The impact of a revision arthroplasty network on patient outcomes. Bone Joint J. 2023;105-b(6):641-8.

13. Feldman ZM, Zheng X, Mao J, Sumpio BJ, Mohebali J, Chang DC, et al. Longer patient travel distance is associated with increased non-index readmission after complex aortic surgery. Journal of Vascular Surgery. 2023;77(6):1607-17.e7.

14. Saunders RS, Fernandes-Taylor S, Kind AJ, Engelbert TL, Greenberg CC, Smith MA, et al. Rehospitalization to primary versus different facilities following abdominal aortic aneurysm repair. J Vasc Surg. 2014;59(6):1502-10, 10.e1-2.

15. Sinha S, Peach G, Poloniecki JD, Thompson MM, Holt PJ. Studies using English administrative data (Hospital Episode Statistics) to assess health-care outcomes--systematic review and recommendations for reporting. Eur J Public Health. 2013;23(1):86-92.

BMJ Open

16. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344-9.

NHS Digital. Hospital Episode Statistics(HES) Analysis Guide. London, UK: NHS Digital;
 2018.

18. Information Standard Board for Health and Social Care. Anonymisation Standard for Publishing Health and Social Care Data Specification. London, UK: NHS Digital; 2013.

19. NHS Digital. Delivering linked datasets to support health and care delivery and research. London, UK: NHS Digital; 2018.

20. Department for Transport. Journey Time Statistics, notes and definitions. London, UK: Department for Transport; 2019.

21. Kalson NS, Mathews JA, Miles J, Bloch BV, Price AJ, Phillips JRA, et al. Provision of revision knee surgery and calculation of the effect of a network service reconfiguration: An analysis from the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. Knee. 2020;27(5):1593-600.

22. Deere K, Whitehouse MR, Kunutsor SK, Sayers A, Price AJ, Mason J, Blom AW. How long do revised and multiply revised knee replacements last? An analysis of the National Joint Registry. Lancet Rheumatol. 2021;3(6):e438-e46.

23. Sabah SA, Knight R, Alvand A, Palmer AJR, Middleton R, Abram SGF, et al. Patient-Relevant Outcomes Following First Revision Total Knee Arthroplasty, by Diagnosis: An Analysis of Implant Survivorship, Mortality, Serious Medical Complications, and Patient-Reported Outcome Measures Utilizing the National Joint Registry Data Set. JBJS. 2023;105(20):1611-21.

24. Kamath CC, O'Byrne TJ, Lewallen DG, Berry DJ, Maradit Kremers H. Neighborhood-Level Socioeconomic Deprivation, Rurality, and Long-Term Outcomes of Patients Undergoing Total Joint Arthroplasty: Analysis from a Large, Tertiary Care Hospital. Mayo Clin Proc Innov Qual Outcomes. 2022;6(4):337-46. Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

25. Lindberg-Larsen M, Jørgensen CC, Bæk Hansen T, Solgaard S, Odgaard A, Kehlet H.
Re-admissions, re-operations and length of stay in hospital after aseptic revision knee
replacement in Denmark: a two-year nationwide study. Bone Joint J. 2014;96-b(12):1649-56.
26. Hardy F, Heyl J, Tucker K, Hopper A, Marchã MJ, Briggs TWR, et al. Data consistency

in the English Hospital Episodes Statistics database. BMJ Health Care Inform. 2022;29(1). 27. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new

framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021;374:n2061.

28. Farrow L, Kennedy JW, Yapp L, Harding T, Walmsley P. Provision of revision knee arthroplasty services across Scotland: A national audit. Knee. 2023;42:312-9.

29. Paterson C, Baarts C, Launsø L, Verhoef MJ. Evaluating complex health interventions: a critical analysis of the 'outcomes' concept. BMC Complement Altern Med. 2009;9:18.

30. Jackson KL, Glasgow RE, Mone MC, Sheng X, Mulvihill SJ, Scaife CL. Does travel distance influence length of stay in elective pancreatic surgery? HPB (Oxford). 2014;16(6):543-9.

31. Cloyd JM, Huang L, Ma Y, Rhoads KF. Predictors of readmission to non-index hospitals after colorectal surgery. The American Journal of Surgery. 2017;213(1):18-23.

32. Kugler CM, De Santis KK, Rombey T, Goossen K, Breuing J, Könsgen N, et al. Perspective of potential patients on the hospital volume-outcome relationship and the

minimum volume threshold for total knee arthroplasty: a qualitative focus group and interview study. BMC Health Serv Res. 2021;21(1):633.

33. Mathews JA, Kalson NS, Tarrant PM, Toms AD. Top ten research priorities for problematic knee arthroplasty. Bone Joint J. 2020;102-b(9):1176-82.

for perteries only



3
4
5
6
7
, 8
0
9
10
11
12
13
14
15
16
17
18
10
20
20
∠ I 22
22
23
24
25
26
27
28
29
30
31
37
22 22
22
34 25
35
36
37
38
39
40
41
42
43
44
15
75 16
40
4/
48
49
50
51
52
53
54
55
56
57
50
50
59
60

Supplementary material S1 – OPCS-4 code criteria used for Hospital Episode
Statistics data extraction
Code Code description
OPCS-4 codes for knee revision procedures

O180	Conversion from previous hybrid prosthetic replacement of knee joint
	using cement
O182	Conversion to hybrid prosthetic replacement of knee joint using

- cement
- O183 Revision of hybrid prosthetic replacement of knee joint using cement
- O184 Attention to hybrid prosthetic replacement of knee joint using cement
- W400 Conversion from previous cemented total prosthetic replacement of knee joint
- W402 Conversion to total prosthetic replacement of knee joint using cement
- W403 Revision of total prosthetic replacement of knee joint using cement
- W404 Revision of one component of total prosthetic replacement of knee joint using cement
- W410 Conversion from previous uncemented total prosthetic replacement of knee joint
- W412 Conversion to total prosthetic replacement of knee joint not using cement
- W413 Revision of total prosthetic replacement of knee joint not using cement
- W414 Revision of one component of total prosthetic replacement of knee joint not using cement
 - W420 Conversion from previous total prosthetic replacement of knee joint NEC
- W422 Conversion to total prosthetic replacement of knee joint NEC

W423	Revision of total prosthetic replacement of knee joint NEC
W424*	Attention to total prosthetic replacement of knee joint NEC
W425	Revision of one component of total prosthetic replacement of knee joint NEC
W522†	Conversion to prosthetic replacement of articulation of bone using cement NEC
W523†	Revision of prosthetic replacement of articulation of bone using cement NEC
W532†	Conversion to prosthetic replacement of articulation of bone not usin cement NEC
W533†	Revision of prosthetic replacement of articulation of bone not using cement NEC
W542†	Conversion to prosthetic replacement of articulation of bone NEC
W543†	Revision of prosthetic replacement of articulation of bone NEC
W544*†	Attention to prosthetic replacement of articulation of bone NEC
W553†	Conversion to prosthetic interposition arthroplasty of joint
W564†	Conversion to interposition arthroplasty of joint NEC
W574†	Conversion to excision arthroplasty of joint
W582†	Revision of resurfacing arthroplasty of joint
W603†	Conversion to arthrodesis and extra-articular bone graft NEC
W613†	Conversion to arthrodesis and articular bone graft NEC
W641†	Conversion to arthrodesis and internal fixation NEC
W642†	Conversion to arthrodesis and external fixation NEC
OPCS-4 c	codes for laterality
7941	Bilateral

3
4
5
6
7
/ 0
8
9
10
11
12
13
14
15
16
17
10
18
19
20
21
22
23
24
25
25
20
27
28
29
30
31
32
33
34
35
26
20
37
38
39
40
41
42
43
44
45
75 76
40
4/
48
49
50
51
52
53
54
54
22
56
57
58
59
60

1 2

Z942	Left-sided					
Z943	Right-sided					
ICD-10 codes for Infection						
T845	Infection and inflammatory reaction due to internal joint prosthesis					
T846	Infection and inflammatory reaction due to internal fixation device [any site]					
T847	Infection and inflammatory reaction due to other internal orthopaedic prosthetic devices, implants and grafts					
T814	Infection following a procedure, not elsewhere classified					
ICD-10 codes for fracture						
M966	Fracture of bone following insertion of orthopaedic implant, joint					
	prosthesis or bone plate					
ICD-10 cc	odes for mechanical complications					
T840	Mechanical complication of internal joint prosthesis					
T841	Mechanical complication of internal fixation device of bones of limb					
T842	Mechanical complication of internal fixation device of other bones					
T843	Mechanical complication of other bone devices, implants and grafts					
T844	Mechanical complication of other internal orthopaedic devices,					
	imnplants and grafts					
	ICD-10 codes for osteoarthritis/arthrosis					
ICD-10 cc	odes for osteoarthritis/arthrosis					
ICD-10 cc M15-	odes for osteoarthritis/arthrosis Polyarthrosis					
ICD-10 cc M15- M17-	odes for osteoarthritis/arthrosis Polyarthrosis Gonarthrosis					
ICD-10 cc M15- M17- M19-	odes for osteoarthritis/arthrosis Polyarthrosis Gonarthrosis Other arthrosis					

Classification of Diseases and Related Health Problems, tenth revision. * Where

OPCS-4 codes Y032 (renewal of prosthesis in organ NOC) or Y037 (removal of prosthesis from organ NOC) were also used. † Where OPCS-4 codes O132 (knee NEC) or Z765 (lower end of femur NEC) or Z774 (upper end of tibia NEC) or Z787 (patella) or Z844 (patellofemoral joint) or Z845 (tibiofemoral joint) or Z846 (knee joint) or Z851 (upper tibiofibular joint) were used to identify knee as the body site.

for oper teries only

Supplementary material S2 – Flow of patient inclusion/exclusions



####Start####

1

3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
3/	
38	
39	
40	
41	
42	
45 11	
+4 ⊿5	
75 76	
40 47	
47 48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	

#Travel Times and Perioperative Outcomes in Revision Knee Replacement ####Preparation of Data#### #load HES data RTKA2023 <- read.csv("~/Desktop/RTKA 06-09-23 CSV.csv") RTKA2023 <- read.csv("/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis /RTKA 06-09-23 CSV.csv") #table only shows first 50 columns but we know there are 51 columns. Write this generic code to change preferences rstudioapi::writeRStudioPreference("data viewer max columns", 1000L) #Find number of incomplete cases in the data missing_data <- colSums(is.na(RTKA2023))</pre> print(missing_data) #There are 4 entried with missing data only in the age group #check how many incomplete entries in age of patient column sum(!complete.cases(RTKA2023\$age_of_patient)) #In case of missing values there are only 4 for age of patient #Can use imputation but given small number decision to remove #What is the mean age of the patients mean(RTKA2023\$age of patient, na.rm = TRUE) #mean age excluding missing values is 70 summary(RTKA2023\$age of patient, na.rm = TRUE) #Check age is normally distributed hist(RTKA2023\$age_of_patient) #we must remove the missing data by coding it NA first

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

RTKA2023\$age of patient[RTKA2023\$age of patient ==""] <- NA **#Remove NA rows** RTKA2023 <- RTKA2023[!is.na(RTKA2023\$age of patient),] #Now check number of missing values sum(!complete.cases(RTKA2023\$age of patient)) #Now states 0 missing values #There are other missing values for IMD decile ##In fact there are 690 IMD score missing values sum(!complete.cases(RTKA2023\$IMD_score)) hist(RTKA2023\$IMD score) #IMD score is non normally distributed summary(RTKA2023\$IMD_score, na.rm = TURE) #Median IMD score is 15. #Use imputation to impute median for missing value RTKA2023\$IMD score[is.na(RTKA2023\$IMD score)] <- 15 #Check imputation complete sum(!complete.cases(RTKA2023\$IMD score)) #Now showing 0 missing values #Next attach IMD decile number 6 to the missing values. As a score of 15 equates to the 6th decile RTKA2023\$IMD decile[is.na(RTKA2023\$IMD decile)] <- 6 #Check duplicate entry spells

duplicates <- RTKA2023[duplicated(RTKA2023),]

print(duplicates)

duplicated(RTKA2023\$P_Spell_ID, fromLast = TRUE)

#No duplicates in data

#Frequencies of revisions by volume

as.numeric(RTKA2023\$TV12mo)

#frequencies of revisions by trust volume table(RTKA2023\$TVcat)

#Proportions by trust volume

prop.table(table(RTKA2023\$TVcat))

#Some entried are blank but are read as real values and not missing data#The table between age and sex shows three variables here#The dataset contains non standard missing values that are not recognised as NA#Replace empty strings with NA

RTKA2023[RTKA2023 == ""] <- NA

#Check this has registered

missing_data <- colSums(is.na(RTKA2023))
print(missing_data)</pre>

#Then remove IMD quintile with NA in rows as only 132 missing #Remove this column

RTKA2023\$IMD_quintile <- NULL

#Column with LSOA_2011_Code has 171 missing. To look at travel times you need to remove these rows

RTKA2023 <- RTKA2023[!is.na(RTKA2023\$LSOA_2011_Code),]

missing_data <- colSums(is.na(RTKA2023))

print(missing_data) #Load Travel times data TRAVELTIMES <- read.csv("~/Desktop/Drive time and Miles reference file.csv") LSOAREF <- read.csv("~/Desktop/LSOA Matrix.csv") #Join data but The data is too big so we need to do this using SQL install.packages("RSQLite") library(RSQLite) con <- dbConnect(RSQLite::SQLite(), dbname = "mydatabase1.db") dbWriteTable(con, "times", TRAVELTIMES) dbWriteTable(con, "Isoa", LSOAREF) query <- " Select * **FROM times** JOIN Isoa ON times.LSOAName = Isoa.LSOA11NM" result <- dbGetQuery(con, query) **#Write Dataframes** write.csv(result, "~/Desktop/JOINLSOATRAVEL.csv") result<- read.csv("~/Desktop/JOINLSOATRAVEL.csv") #####Now join this data to your revisions spreadsheet using key identifiers LSOA and Organisation site code con <- dbConnect(RSQLite::SQLite(),</pre> dbname = "mydatabase1.db") dbWriteTable(con, "revisions2", RTKA2023) dbWriteTable(con, "travel2", result) query <- " Select * FROM revisions2 JOIN travel2 ON revisions2.LSOA 2011 Code = travel2.LSOA11CD AND revisions2.Sitecode = travel2.ProviderSiteCode"

BMJ Open

result1 <- dbGet	tQuery(con, query)		
write.csv(result	1, "~/Desktop/REVISION	NSTRAVELTIMES.csv")	
result2<- read.c	sv("~/Desktop/REVISIO	NSTRAVELTIMES.csv")
#Check your dat	ta for missing values		
missing_data <- print(missing_d	colSums(is.na(result1)) ata)		
####Prepare Ou	itcomes, Exposure varia	able and co-variates #	###
#Set up outcom	es		
#Replace NA's ii	n the Read columns wit	h N	
result1\$Read30 result1\$Read90	<- ifelse(is.na(result1\$ <- ifelse(is.na(result1\$	Read30), 'N', result1\$ Read90), 'N', result1\$	Read30) Read90)
result1\$Read30 #readmission fo result1\$Read90	days <- ifelse(result1\$R r 90 days days <- ifelse(result1\$R	ead30 == "Y", 1, 0) ead90 == "Y", 1, 0)	
#Set up your co	-variates		
result1\$HFRS_B result1\$HFRS_B	and = as.factor(result1 and = relevel(result1\$F	\$HFRS_Band) IFRS_Band, ref = 'Non	e')
result1\$POD = a result1\$POD = r	is.factor(result1\$POD) elevel(result1\$POD, ref	⁼ 'EL')	
table(result1\$P(OD)		
#Sensitivity ana	lysis for only aseptic ca	ses	
result2 <- subse	t(result1, infection == 0))	
#Subset the dat include volume	a to focus on teritary co categories D,E & F	entres only determine	d by volume >59. Therefore
Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

#Trust volume was categorised as < 20, 20-39, 40-59, 60-79, 80-99 and \geq 100 procedures in the previous year. These categories were chosen to ensure that there were more than ten trusts/surgeons represented in each category and that the categorisations were meaningful and consistent.

traveltimesrev <- subset(result2, TVcat == "D" | TVcat == "E" | TVcat == "F")

#≥70 a year BASK recommendations for Major Revision Centres

result2\$MRC <- ifelse(result2\$TV12mo > 70, 1, 0)

traveltimesrev <- subset(result2, MRC == 1)

#Create travel time quintile variable

quintiles <- quantile(traveltimesrev\$DistanceMiles, probs = seq(0,1,0.2), na.rm=TRUE)

traveltimesrev\$distancequintile <- cut(traveltimesrev\$DistanceMiles, breaks = quintiles, labels = c("Q1", "Q2", "Q3", "Q4", "Q5"), include.lowest = TRUE)

#Add new outcome variable LOS>median

summary(traveltimesrev\$Spell_Los)
#Spell length of stay median is 5 days

traveltimesrev\$LongLOS <- ifelse(traveltimesrev\$Spell_Los >5, 1,0)

#Add IMD quintiles to look at this association with the outcome

quintiles <- quantile(traveltimesrev\$IMD_score, probs = seq(0,1,0.2), na.rm=TRUE)</pre>

traveltimesrev\$IMD_quintile <- cut(traveltimesrev\$IMD_score, breaks = quintiles, labels = c("Q1", "Q2", "Q3", "Q4", "Q5"), include.lowest = TRUE)

####Save final dataset
write.csv(traveltimesrev, "~/Desktop/REVISIONSTRAVELTIMESFINAL.csv")

###Load final dataset

traveltimesrev <- read.csv("/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/REVISIONSTRAVELTIMESFINAL.csv")

####Descriptive Statistics####

BMJ Open

2			
3	#Describe raw count statistics based on stratified travel quintile		
4			
5	table/traveltimeerav(cdistance quintile)		
6	(abie(traveitimesrevșuistancequintile)		
/			
8	summary (traveltimes rev \$ Distance Miles)		
9			
10	table(traveltimesrev\$ageband, traveltimesrev\$distancequintile)		
12			
13			
14	summary/travoltimesrov(cdistancequintile)		
15	summary(traveitimesrev;suistancequintile)		
16			
17	table(traveltimesrev\$Fractue)		
18	table(traveltimesrev\$Mechanical.complication)		
19	table(traveltimesrev\$OA)		
20	table(traveltimesrev\$RTKA_nonspecific)		
21	table(traveltimesrev\$conversion to TKA)		
22	table/traveltimesrev\$conversion.co.nka/		
23	table(traveitimesrev\$one_component)		
24	table(traveltimesrev\$Attention.to)		
25			
20	table(traveltimesrev\$distancequintile, traveltimesrev\$Read30days)		
27			
29	table(traveltimesrev\$Read30days)		
30			
31	table/traveltimecrov/Drovider Name)		
32			
33			
34	table(traveltimesrev\$distancequintile, traveltimesrev\$Read90days)		
35			
36	table(traveltimesrev\$distancequintile, traveltimesrev\$Mort90days)		
37			
38	table(traveltimesrev\$distancequintile_traveltimesrev\$rev1vr)		
39			
40 41	table (the calification of distance and initial three all time a security in a 1.00)		
42	table(traveitimesrev\$distancequintile, traveitimesrev\$LongLOS)		
43			
44	#Demographics and Clinical Characteristics		
45			
46	table(traveltimesrev\$distanceguintile, traveltimesrev\$IMD_guintile)		
47			
48	table/traveltimesrov\$distancequintile_traveltimesrov\$sov}		
49	table(traventimesrevoustancequintile, traventimesrevosex)		
50			
51	table(traveltimesrev\$distancequintile, traveltimesrev\$Mechanical.complication)		
52			
53	table(traveltimesrev\$distancequintile, traveltimesrev\$Fractue)		
54 55			
56	table(traveltimesrev\$distancequintile, traveltimesrev\$OA)		
57			
58	table/traveltimecrov&dictancequintile_traveltimecrov&UEDS_Band		
59	table(traveitimesrevșuistancequintile, traveitimesrevșitro_banu)		
60			

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

> table(traveltimesrev\$distancequintile, traveltimesrev\$CVcat) ####Correlations#### #Find out if IMD score or Age as continous are associated with Travel distance #Look at median age and IMD in each of the travel distance quintiles first new1 <- subset(traveltimesrev, distancequintile == "Q1")</pre> new2 <- subset(traveltimesrev, distancequintile == "Q2") new3 <- subset(traveltimesrev, distancequintile == "Q3") new4 <- subset(traveltimesrev, distancequintile == "Q4") new5 <- subset(traveltimesrev, distancequintile == "Q5") #Calculate median age for each travel quintile summary(new1\$age_of_patient) summary(new2\$age of patient) summary(new3\$age_of_patient) summary(new4\$age of patient) summary(new5\$age_of_patient) boxplot(traveltimesrev\$age of patient ~ traveltimesrev\$distancequintile, xlab = "Travel Distance Quintile", ylab = "Age (years)") #Calculate median IMD score for each travel quintile summary(new1\$IMD score) summary(new2\$IMD score) summary(new3\$IMD_score) summary(new4\$IMD score) summary(new5\$IMD score) boxplot(traveltimesrev\$IMD score ~ traveltimesrev\$distancequintile, xlab = "Travel Distance Quintile", ylab = "IMD Score") #Next do a Spearman's rank correlation between travel distance and age, and then for travel distance and IMD score cor.test(traveltimesrev\$age of patient, traveltimesrev\$DistanceMiles, method="spearman")

BMJ Open

3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
52	
50	
72	
00	

cor.test(traveltimesrev\$IMD score, traveltimesrev\$DistanceMiles, method="spearman") #Find the median travel time for patients in Q5 travel quintile new <- subset(traveltimesrev, distancequintile == "Q5") summary(new\$DistanceMiles) #Find median travel distance for each travel quintile new1 <- subset(traveltimesrev, distancequintile == "Q1") new2 <- subset(traveltimesrev, distancequintile == "Q2")</pre> new3 <- subset(traveltimesrev, distancequintile == "Q3") new4 <- subset(traveltimesrev, distancequintile == "Q4") new5 <- subset(traveltimesrev, distancequintile == "Q5")</pre> summary(new1\$DistanceMiles) summary(new5\$DistanceMiles) #Repeat for other distance quintiles ####Modelling#### **#Logistic Regression** #Primary outcome variable binary admitted within 30 days or not model.log<-glm(Read30days ~ distancequintile + IMD quintile + sex + ageband + Mechanical.complication + Fractue + OA + HFRS Band + CVcat, data=traveltimesrev, family = "binomial") summary(model.log) exp(cbind(OR = coef(model.log), confint(model.log, level = 0.95))) install.packages("MASS") library("MASS") #Mass is loaded in other packages such as Imertest OR CI <- round(exp(cbind(coef(model.log),</pre>

confint(model.log))), digits = 3)

```
result_table <- data.frame(
  Coefficient = coef(model.log),
  P_Value = summary(model.log)$coefficients[, "Pr(>|z|)"]
)
```

```
write.csv(result_table, "~/Desktop/Sensitivty MORT.csv")
```

#Plot graph

#this creates a matrix, we now need to convert into a dataframe and change column names

df <- as.data.frame(OR_CI)

#Remove intercept row the first row

df = df[-1,]

#add covariate column

df\$covariate <- c('Distance quintile 2 (ref: Q1)', 'Distance quintile 3 (ref: Q1)', 'Distance quintile 4 (ref: Q1)', 'Distance quintile 5 (ref: Q1)', 'IMD_quintileQ2 (ref:Q1)', 'IMD_quintileQ3 (ref:Q1)', 'IMD_quintileQ4 (ref:Q1)', 'IMD_quintileQ5 (ref:Q1)', 'Male vs Female', '60-64', '65-69', '70-74', '75-79', '>=80', 'Mechanical failure vs no failure', 'Fracture vs no fracture', 'Progressive OA vs no OA', 'HFRS_Band Mild (ref: None)', 'HFRS_Band Moderate (ref:None)', 'HFRS_Band Severe (ref:None)', 'Surgeon annual volume 5-9 (ref 0-4)', 'Surgeon annual volume 10-14 (ref 0-4)', 'Surgeon annual volume 15-19 (ref 0-4)', 'Surgeon annual volume 20-24 (ref 0-4)', 'Surgeon annual volume >=25 (ref 0-4)')

#Save dataframe to desktop for analysis write up

write.csv(df, "~/Desktop/sensitivty MORT Log.csv")

```
ggplot(data=df, aes(y = df$covariate, x = df$V1, xmin=df$`2.5 %`,xmax=df$`97.5 %`))+
geom_point()+
geom_errorbarh(height=.1)+
geom_vline(xintercept = 1)+
xlab("Odds Ratio")+
ylab("Exposure & Co-variates")+
ggtitle("Odds for mortality within 90 days")
```

BMJ Open

2	
3	
4	#Cours Odde ratiols and OFO/ confidence intervals as now detaframe
5	#Save Odds ratio's and 95% confidence intervals as new dataframe
6	
7	coefficients table <- as.data.frame(exp(cbind(OR = coef(model.log), confint(model.log, level
8	- 0.05))))
9	- 0.95////
10	
11	write.csv(coefficients_table, "~/Desktop/MultivariableLogisticRegression.csv")
17	
12	
13	
14	
15	#No statistical difference in 30 day readmission rates between different quintiles
16	
17	
18	#Risk of LOS>median
19	model.log<-glm(LongLOS ~ distancequintile + IMD_quintile + sex + ageband +
20	Mechanical complication + Fractue + OA + HERS_Band + CVcat_data=traveltimesrev_family
21	- "binomial")
22	
23	summary(model.log)
24	exp(cbind(OR = coef(model.log), confint(model.log, level = 0.95)))
25	#No statistical difference for LOS between quintiles adjusted
26	#NO statistical difference for LOS between quintiles adjusted
27	
28	
29	#Mortality at 90 days
30	model log < glm/Mert00deve & distance quintile + IMD, quintile + cov + exchand +
31	model.iog<-gim(ivior1900ays distancequintile + iviD_quintile + sex + ageband +
37	Mechanical.complication + Fractue + OA + HFRS_Band + CVcat, data=traveltimesrev, family
33	= "binomial")
27	summary(model log)
24 25	$\frac{1}{2} = \frac{1}{2} \left(\frac{1}{2} + 1$
35	exp(cbind(OR = coet(model.log), contint(model.log, level = 0.95)))
30	
37	#No difference for mortality at 90 days
38	
39	
40	#Testing for model fit
41	
42	null <- glm(rev1vr ~ 1_data = traveltimesrev_family = "hinomial")
43	Har
44	#Of
45	null <- glm(Mort90days ~ 1, data = traveltimesrev, family = "binomial")
46	#or
47	null <_ alm/Read30days ~ 1_data - traveltimesrey family - "hinomial")
48	indi <- gini(headbodays 1, data - traventinesrev, faining - binomiar)
49	#or
50	null <- glm(Read90days ~ 1, data = traveltimesrev, family = "binomial")
51	#or
52	null c. alm/Madianlas X.1. data - travaltimagray, family - "hinamial")
52	nun <- gim(ivieulanios 1, uala = traveilimesrev, lanniy = binomiai)
54	
55	anova(model.log, null, test = "Chisq")
55	
50	LDT a mandal la sérvul de vien en una del la série de vien en
5/	LKT <- Model.log\$null.deviance - model.log\$deviance
58	print(LRT)
59	
60	

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
22 22
23 24
24 25
20 26
20 27
27 20
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 2

#This shows a non significant X2 statistics which shows the model has a good fit

#The best way to check for collinearity is using VIF

#variance inflation factor (or VIF), which measures how much the variance of a regression coefficient is inflated due to multicollinearity in the model.

install.packages("car") library(car)

plot(model.log)

install.packages("carData") library(carData)

#You need to run this code for each model used in Logistic Regression

vif(model.log)

ols_vif_tol(model.log)

#For each outcome model (logistic regression) VIF is <3 therefore

#None of the VIF exceeds 5 so we can assume there is no evidence of strong multicollinearity

shapiro.test(rstandard(model.log))

#Shapiro wilcox test shows no evidence of multicollinearity, very low p value so we can reject the null hypothesis of normality

#Calculate pseudo R squared values at assess model fit

II.full<-logLik(model.log)
II.null<-logLik(null)
n<-length(model.log\$residuals)</pre>

McFadden Test as.numeric(1-(II.full/II.null))

#Evidence showing good model fit for all models

#####Sensitivty analysis Drive Distances#### #Create drive time quintile variable

BMJ Open

qui na.	intiles <- quantile(traveltimesrev\$OffPeakDriveDistanceMiles, probs = seq(0,1,0.2), rm=TRUE)
trav bre	veltimesrev\$drivedistancequintile <- cut(traveltimesrev\$OffPeakDriveDistanceMiles, eaks = quintiles, labels = c("Q1", "Q2", "Q3", "Q4", "Q5"), include.lowest = TRUE)
#Fii	nd median off peak distance by quintile
nev	w1 <- subset(traveltimesrev, drivedistancequintile == "Q1")
nev	w2 <- subset(traveltimesrev, drivedistancequintile == "Q2")
nev	w3 <- subset(traveltimesrev, drivedistancequintile == "Q3")
nev	w4 <- subset(traveltimesrev, drivedistancequintile == "Q4")
nev	w5 <- subset(traveltimesrev, drivedistancequintile == "Q5")
sun sun sun	nmary(traveltimesrev\$OffPeakDriveDistanceMiles) nmary(new1\$OffPeakDriveDistanceMiles) nmary(new5\$OffPeakDriveDistanceMiles)
#Th	ne calculate the primary and secondary outcomes again
#Lc	ogistic Regression
#Pr	rimary outcome variable binary admitted within 30 days or not
mo Me = "l	odel.log<-glm(Read30days ~ drivedistancequintile + IMD_quintile + sex + ageband + echanical.complication + Fractue + OA + HFRS_Band + CVcat, data=traveltimesrev, family binomial")
sun exp	nmary(model.log) p(cbind(OR = coef(model.log), confint(model.log, level = 0.95)))
#Ri	isk of LOS>median
mo Me = "l	odel.log<-glm(LongLOS ~ drivedistancequintile + IMD_quintile + sex + ageband + echanical.complication + Fractue + OA + HFRS_Band + CVcat, data=traveltimesrev, family binomial")
sun exp #No	nmary(model.log) o(cbind(OR = coef(model.log), confint(model.log, level = 0.95))) o statistical difference for LOS between quintiles adjusted
#M mo Me = "l	lortality at 90 days odel.log<-glm(Mort90days ~ drivedistancequintile + IMD_quintile + sex + ageband + ochanical.complication + Fractue + OA + HFRS_Band + CVcat, data=traveltimesrev, family binomial")

exp(cbind(OR = coef(model.log), confint(model.log, level = 0.95))) #No statistical difference for LOS between quintiles adjusted
<pre>#Mortality at 90 days model.log<-glm(Mort90days ~ timequintile + IMD_quintile + sex + ageband + Mechanical.complication + Fractue + OA + HFRS_Band + CVcat, data=traveltimesrev, family = "binomial") summary(model.log) exp(cbind(OR = coef(model.log), confint(model.log, level = 0.95)))</pre>
####Sup Material S4 Crude Rates #### #Supplementary Tables travel as continous variable
#Plot crude rates of 30 day readmission and road distances with off peak journeys in mind
<pre># Calculate failure rates by surgical unit hospital_failure_rates <- traveltimesrev %>% group_by(OffPeakDriveDistanceMiles) %>% summarise(total_surgeries = n(), total_failures = sum(Read30days, na.rm = TRUE), failure_rate = total_failures / total_surgeries)</pre>
Remove any rows with NA values in relevant columns before fitting hospital_failure_rates_clean <- hospital_failure_rates %>% filter(!is.na(OffPeakDriveDistanceMiles), !is.na(failure_rate))
Fit the LOESS model to the cleaned data loess_fit <- loess(failure_rate ~ OffPeakDriveDistanceMiles, data = hospital_failure_rates_clean)
Make predictions on the cleaned data predictions <- predict(loess_fit, newdata = hospital_failure_rates_clean, se = TRUE)
Add the predictions back to the cleaned dataset hospital_failure_rates_clean\$fit <- predictions\$fit hospital_failure_rates_clean\$se <- predictions\$se.fit
<pre>ggplot(hospital_failure_rates_clean, aes(x = OffPeakDriveDistanceMiles, y = failure_rate)) + geom_point(alpha = 0.5) + geom_line(aes(y = fit), color = "blue") + # Add the fitted line geom_ribbon(aes(ymin = pmax(0, fit - 1.96 * se), ymax = fit + 1.96 * se), alpha = 0.2) + # 95% CI with lower bound constrained to 0</pre>

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

```
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

```
labs(
  x = "Off Peak Drive Distance (Miles)",
  y = "Readmission within 30 days",
  title = "LOESS Fit: Re-admission within 30 days by travel distance"
 )+
 scale y continuous(labels = scales::percent format(), limits = c(0, NA)) +
 scale_x_continuous(limits = c(0, max(hospital_failure_rates$OffPeakDriveDistanceMiles)))
#Crude rates and travel distance as crow flies
# Calculate failure rates by surgical unit
hospital failure rates <- traveltimesrev %>%
 group by(DistanceMiles) %>%
 summarise(
  total surgeries = n(),
  total failures = sum(Read30days, na.rm = TRUE),
  failure_rate = total_failures / total_surgeries
 )
# Remove any rows with NA values in relevant columns before fitting
hospital failure rates clean <- hospital failure rates %>%
filter(!is.na(DistanceMiles), !is.na(failure_rate))
# Fit the LOESS model to the cleaned data
loess_fit <- loess(failure_rate ~ DistanceMiles, data = hospital_failure_rates_clean)
# Make predictions on the cleaned data
predictions <- predict(loess fit, newdata = hospital failure rates clean, se = TRUE)
# Add the predictions back to the cleaned dataset
hospital failure rates clean$fit <- predictions$fit
hospital failure rates clean$se <- predictions$se.fit
ggplot(hospital failure rates clean, aes(x = DistanceMiles, y = failure rate)) +
 geom point(alpha = 0.5) +
 geom line(aes(y = fit), color = "blue") + # Add the fitted line
 geom_ribbon(aes(ymin = pmax(0, fit - 1.96 * se), ymax = fit + 1.96 * se), alpha = 0.2) + #
95% CI with lower bound constrained to 0
 labs(
  x = "As Crow Flies Travel Distance (Miles)",
  y = "Readmission within 30 days",
  title = "LOESS Fit: Re-admission within 30 days by travel distance"
 )+
 scale y continuous(labels = scales::percent format(), limits = c(0, NA)) +
```

scale_x_continuous(limits = c(0, max(hospital_failure_rates\$DistanceMiles)))
#Crude rates peak drive time and 30 day re-admission
<pre># Calculate failure rates by surgical unit hospital_failure_rates <- traveltimesrev %>% group_by(PeakDriveTime) %>% summarise(total_surgeries = n(), total_failures = sum(Read30days, na.rm = TRUE), failure_rate = total_failures / total_surgeries)</pre>
<pre># Remove any rows with NA values in relevant columns before fitting hospital_failure_rates_clean <- hospital_failure_rates %>% filter(!is.na(PeakDriveTime), !is.na(failure_rate))</pre>
Fit the LOESS model to the cleaned data
loess_fit <- loess(failure_rate ~ PeakDriveTime, data = hospital_failure_rates_clean)
<pre># Make predictions on the cleaned data predictions <- predict(loess_fit, newdata = hospital_failure_rates_clean, se = TRUE) # Add the predictions back to the cleaned dataset hospital_failure_rates_clean\$fit <- predictions\$fit hospital_failure_rates_clean\$se <- predictions\$se.fit</pre>
<pre>ggplot(hospital_failure_rates_clean, aes(x = PeakDriveTime, y = failure_rate)) + geom_point(alpha = 0.5) + geom_line(aes(y = fit), color = "blue") + # Add the fitted line geom_ribbon(aes(ymin = pmax(0, fit - 1.96 * se), ymax = fit + 1.96 * se), alpha = 0.2) + # 95% Cl with lower bound constrained to 0 labs(x = "Peak Drive Time (Minutes)", y = "Readmission within 30 days", title = "LOESS Fit: Re-admission within 30 days by travel times") + scale_y_continuous(labels = scales::percent_format(), limits = c(0, NA)) + scale_x_continuous(limits = c(0, max(hospital_failure_rates\$PeakDriveTime)))</pre>

:	####Supp Material S4 Logistic Regression#### #Logisic Regression Model Distance Miles Primary Outcome
:	#####Model 1 unadjusted
	model <- glm(Read30days ~ DistanceMiles, family = binomial(link = "logit"), data = traveltimesrev)
:	summary(model)
:	#p value 0.763, -0.002439 coef, AIC 797.95
:	#null models #re-revision at 2 yrs null <- glm(Read30days ~ 1, data = traveltimesrev, family = "binomial")
	II.full <- logLik(model) II.null <- logLik(null) n<- length(model\$residuals) as.numeric(1-(II.full/II.null))
:	#r squared 0.000117
:	#Patient Factors
	model<-glm(Read30days ~ DistanceMiles + IMD_quintile + sex + ageband + Mechanical.complication + Fractue + OA + HFRS_Band, data=traveltimesrev, family = "binomial") summary(model)
:	#AIC 796.67, p value 0.976. r squared 0.04
:	#Surgeon Factors
:	model<-glm(Read30days ~ DistanceMiles + IMD_quintile + sex + ageband + Mechanical.complication + Fractue + OA + HFRS_Band + CVcat, data=traveltimesrev, family = "binomial") summary(model) exp(cbind(OR = coef(model), confint(model, level = 0.95)))
:	#AIC 797.31, p value 0.912, coef 0.009223, r squared 0.0538
:	#No statistical relationship between as crow flies travel distance and primary outcome

#Is travel distance linear or non linear
#Box Tidwell
model <- glm(Read30days ~ DistanceMiles, family = binomial(link = "logit"), data = traveltimesrev)
coef_summary <- summary(model)
<pre>box_tidwell_Mean_Unit <- coef_summary\$coefficient[2, "Pr(> z)"]</pre>
print(box_tidwell_Mean_Unit)
#p value 0.762, it is not non-linear, no indication to model with splines
#Model as categorical quintiles and assess model fit for comparison
model<-glm(Read30days ~ distancequintile + IMD_quintile + sex + ageband + Mechanical.complication + Fractue + OA + HFRS_Band + CVcat, data=traveltimesrev, family = "binomial") summary(model)
#AIC 794.68, r squared 0.064 (improved model fit)
#Logistic regression travel distance by road
#####Model 1 unadjusted
model <- glm(Read30days ~ OffPeakDriveDistanceMiles, family = binomial(link = "logit"), data = traveltimesrev)
summary(model)
#p value 0.544, -0.003686 coef, AIC 797.66
#null models #re-revision at 2 yrs null <- glm(Read30days ~ 1, data = traveltimesrev, family = "binomial")
II.full <- logLik(model) II.null <- logLik(null) n<- length(model\$residuals) as.numeric(1-(II.full/II.null))

#r squared 0.000475
#Patient Factors
model<-glm(Read30days ~ OffPeakDriveDistanceMiles + IMD_quintile + sex + ageband + Mechanical.complication + Fractue + OA + HFRS_Band, data=traveltimesrev, family = "binomial") summary(model)
#AIC 796.6, p value 0.787. r squared 0.0421, coef -0.00168
#Surgeon Factors
model<-glm(Read30days ~ OffPeakDriveDistanceMiles + IMD_quintile + sex + ageband + Mechanical.complication + Fractue + OA + HFRS_Band + CVcat, data=traveltimesrev, family = "binomial") summary(model)
#AIC 797.3, p value 0.882, coef -0.00093, r squared 0.0538
exp(cbind(OR = coef(model), confint(model, level = 0.95)))
#No statistical relationship between as crow road travel distance and primary outcome
#Is travel distance linear or non linear
#Box Tidwell
model <- glm(Read30days ~ OffPeakDriveDistanceMiles, family = binomial(link = "logit"), data = traveltimesrev)
coef_summary <- summary(model)
box_tidwell_Mean_Unit <- coef_summary\$coefficient[2, "Pr(> z)"]
print(box_tidwell_Mean_Unit)
#p value 0.544, it is not non-linear, no indication to model with splines
#Model as categorical quintiles and assess model fit for comparison
model<-glm(Read30days ~ distancequintile + IMD_quintile + sex + ageband + Mechanical.complication + Fractue + OA + HFRS_Band + CVcat, data=traveltimesrev, family = "binomial") summary(model)

2	
3	
4	
5	
б	#AIC 794.68, r squared 0.064 (improved model fit)
7	
8	
9	
10	
10	
11	
12	
13	#Logistic regression Road Travel Times
14	
15	#####Model 1 upadiusted
16	
17	
18	
19	model <- glm(Read30days ~ PeakDriveTime, family = binomial(link = "logit"), data =
20	trovaltimosrov)
21	travertimesrev)
22	
23	summary(model)
24	
25	
25	# AIC 797.77, p value 0.608, coef -0.002451
20	
27	#null models
20	
29	#re-revision at 2 yrs
30	null <- glm(Read30days ~ 1, data = traveltimesrev, family = "binomial")
31	
32	
33	
34	II.full <- logLik(model)
35	II.null <- logLik(null)
36	n<- length(modelSresiduals)
37	
38	as.numeric(1-(II.full/II.null))
39	
40	#r squared 0.000336
41	
47	
43	#Patient Factors
45 11	
44	model<-glm(Read30days ~ PeakDriveTime + IMD, quintile + sex + agehand +
45	Masha sisaha analisatian a Frankova OA a UEDC David data tara akimana famil
40	Mechanical.complication + Fractue + OA + HFRS_Band, data=traveltimesrev, family =
47	"binomial")
48	summary(model)
49	
50	
51	#AIC 796.65, p value 0.863. r squared 0.042
52	
53	#Surgeon Factors
54	
55	
56	model<-glm(Read30days ~ PeakDriveTime + IMD quintile + sex + ageband +
57	Mechanical complication + Fractue + OA + HERS Band + CVcat data=traveltimesrey family
58	- "binomial")
59	= Dinomial)
60	summary(model)

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

exp(cbind(OR = coef(model), confint(model, level = 0.95))) #AIC 797.32, p value 0.968, coef 0.0002027, r squared 0.0538 #No statistical relationship between drive time and primary outcome #Is travel distance linear or non linear **#Box Tidwell** model <- glm(Read30days ~ PeakDriveTime, family = binomial(link = "logit"), data = traveltimesrev) coef_summary <- summary(model)</pre> box_tidwell_Mean_Unit <- coef_summary\$coefficient[2, "Pr(>|z|)"] print(box tidwell Mean Unit) #p value 0.608, it is not non-linear, no indication to model with splines #Model as categorical quintiles and assess model fit for comparison model<-glm(Read30days ~ timequintile + IMD quintile + sex + ageband + Mechanical.complication + Fractue + OA + HFRS_Band + CVcat, data=traveltimesrev, family = "binomial") summary(model) #AIC 800, r squared 0.058 ####END####

Supplementary material S4 – Relationship between Travel distances and times modelled as a continuous variable with primary outcome (readmission within 30 days)

Crude Rates of 30-day readmission and Straight-Line Travel Distance (Locally Estimated Scatterplot Smoothing Fit used to estimate trends with standard errors)



Crude Rates of 30-day readmission and Travel Distance by Road (Locally Estimated Scatterplot Smoothing Fit used to estimate trends with standard errors)



Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Crude Rates of 30-day readmission and Travel Times by Road (Locally Estimated Scatterplot Smoothing Fit used to estimate trends with standard errors)

LOESS Fit: Re-admission within 30 days by travel times



Multiple Variable Logistic Regression for Travel Distances and Times (continous) and readmission within 30 days.

	Association between Travel Distances and Times (continuous) and Readmission within 30 days		
	Odds Ratio/Coefficient estimate (95% confidence intervals)	p value	R ²
Straight Line Travel Distance	1.00 (0.98 to 1.02)	0.91	5.38%

3	Travel Distance by	1.00 (0.99 to 1.01)	0.88	5.38%
4 5	Road			
5	Peak Road Travel	1.00 (0.99 to 1.01)	0.97	5.38%
7	Times			
8		1		
9				
10				
11				
12				
14				
15				
16				
17				
10				
20				
21				
22				
23 24				
25				
26				
27				
28				
29 30				
31				
32				
33				
34 35				
36				
37				
38				
39 40				
40				
42				
43				
44				
45 46				
47				
48				
49				
50 51				
52				
53				
54				
55				
50 57				
58				
59				
60				

What is the impact of longer patient travel distances and times on perioperative outcomes following revision knee replacement: An analysis using national administrative data from Hospital Episode Statistics

Journal:	BMJ Open
Manuscript ID	bmjopen-2024-085201.R2
Article Type:	Original research
Date Submitted by the Author:	16-Jan-2025
Complete List of Authors:	Matthews, Alexander; University of Exeter Faculty of Health and Life Sciences, Department of Clinical and Biomedical Sciences; Royal Devon and Exeter NHS Foundation Trust, Evans, Jonathan; University of Exeter Medical School, Health Services and Policy Research; Royal Devon and Exeter NHS Foundation Trust, Princess Elizabeth Orthopaedics Centre Evans, Jonathan; University of Exeter, Department of Public Health and Sports Sciences Lamb, Sarah; University of Exeter, College of Medicine and Health Price, Andrew; University of Oxford, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences Gray, William; NHS England and NHS Improvement London Briggs, Tim; NHS England and NHS Improvement London Toms, Andrew ; Princess Elizabeth Orthopaedic Centre
Primary Subject Heading :	Public health
Secondary Subject Heading:	Surgery
Keywords:	Knee < ORTHOPAEDIC & TRAUMA SURGERY, PUBLIC HEALTH, Health Services

SCHOLARONE[™] Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our <u>licence</u>.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which <u>Creative Commons</u> licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

terez oni

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies



Title What is the impact of longer patient travel distances and times on perioperative outcomes following revision knee replacement: An analysis using national administrative data from Hospital Episode Statistics Names, Affiliations, and positions of all authors Alex Matthews1,2,3,4, Jonathan P Evans2,3, Jonathan T Evans2,3, Sarah E Lamb3 Andrew Price4,5, William Gray1, Tim Briggs1,6, Andrew Toms2,3 1. Getting It Right First Time programme, NHS England, London, UK Royal Devon University Healthcare NHS Foundation Trust, Exeter, UK 3. University of Exeter, Exeter, UK 4. Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, UK 5. Nuffield Orthopaedic Centre, Oxford, UK 6. Royal National Orthopaedic Hospital, Stanmore, London, UK

22	Correspondence: Alexander Matthews, Getting It Right First Time programme,			
23	NHS England and NHS Improvement, Wellington House, Waterloo Road. London,			
24	UK, SE1 8UG; telephone: 07401932627 email: alexander.matthews1@nhs.net			
25	Word count: [3316 words]			
26	Short title: Patient Journey Distance and Time for Complex Revision Knee Surgery			
27 28 20	Key words: social determinants of health, revision knee replacement, contextual factors in complex health interventions			
29 30	Twitter handles:			
31	Alex Matthews: @DrAlexMatthews			
32	Tim WR Briggs: @ProfTimBriggs			
33	WilliamKGray1: @WilliamKGray1			
34	GIRFT programme: @NHSGIRFT			
35 36				
37	Structured Abstract			
38 39	Objectives			
40 41	Patients undergoing revision total knee replacement (RevKR) surgery often have			
42	difficulties mobilising and increasingly rely on family support. Evolving practice in			
43	England aims to manage these patients in specialised centres with the intention of			
44	improving outcomes. This practice will result in longer travel distances and times in			
45	this frailer group of patients. We want to examine the types of distances and travel			
46	times patients can be expected to travel for this complex orthopaedic surgery and to			
47	explore concerns of how these impact patient outcomes.			
48	Design			
49 50	Retrospective observational study from the Hospital Episode Statistics. Pooled			
51	multivariable adjusted logistic regression models were used to investigate the			
52	relationship between patient travel distances and times with perioperative outcomes			
53 54	Setting			
	22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54			

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Patients presenting to tertiary referral centres between 1st January 2016 to 31st

December 2019. A tertiary referral centre was defined as a trust performing >49 revisions in the year prior. **Participants** Adult patients undergoing RevKR procedures for any reason between 1st January 2016 to 31st December 2019. Exposure The shortest patient level travel distance and time was calculated using the department of health Journey Time Statistics using TRACC software and Dijkstra's algorithm. Main Outcome Measures The primary outcome is emergency readmission within 30 days. Secondary outcomes are mortality within 90 days and length of inpatient stay. Results 6,880 patients underwent RevKR at 36 tertiary referral centres. There was a weak correlation between social deprivation and travel distance, with patients from the most deprived areas travelling longer distances. Overall, 30-day readmission was not statistically associated longer driving distance (OR 1.00 95% CI 0.99 to 1.02) or peak driving times (OR 1.00 95% CI 0.99 to 1.01). Conclusions There was no association between increasing travel distance and time on perioperative outcomes for RevKR patients.

1		
2		
3	93	
4	94	
5	95	
0 7	96	
/	07	
0	00	Strangthe and limitations of this study
9 10	98	
11 12	99	Our study is the first to describe patient travel distance and time associations
13 14	100	using a large longitudinal dataset.
15 16	101	This data reflects revision knee replacement procedures undertaken across
17 18	102	different geographical areas of England
19 20	103	 Owing to differences in the coverage of Hospital Episode Statistics,
21 22	104	procedures in hospitals outside of England were not included in this analysis
23 24	105	Clinical coding practice within HES is known to vary between trusts but this is
25	106	unlikely to be yary systematically to bias our findings
26 27	100	uninkely to be vary systematically to bias our infollings
28 29	107	 This analysis only reports travel times for patients with access to their own
30 31	108	transport and does not consider times for those patients using public transport
32 33	109	
34 35	110	
36 37	111 112	
38 39	112	
40		
41 42		
43		
44		
45		
40 47		
48		
49		
50		
51		
52		
53		
54 57		
55 56		
57		
58		
59		
60		

113 Introduction

Primary knee replacement is a successful procedure that improves quality of life for the majority of patients.[1] However, at 10 years following a primary knee replacement, about 3.5% of patients will have undergone a revision surgery.[2] The majority of these procedures are carried out due to infection or polyethylene wear of the implant.[3] A failed primary knee replacement represents a life changing transition point where individuals are likely to suffer from pain, reduced mobility as well as dependency on family members. [4] Patients often face multi-step surgery with longer hospital length of stays and higher complication rates [5, 6]

The Getting It Right First Time (GIRFT) programme orthopaedic National Report was published in 2015.[7] A key recommendation was the centralisation of complex orthopaedic surgery, including revision knee surgery, to specialist centres with the aim of improved patient outcomes. Consequently, revision total knee replacement (RevKR) surgery in the England has evolved into a regional network service model.[8] All hospitals performing RevKR form a network in the respective regions. Less specialist hospitals, defined by lower annual case volume thresholds, are encouraged to discuss and sometimes refer their caseload to more specialist centres. Several studies based on large revision hip and knee registries have suggested this model carries a lower failure rate defined by the need for further revision surgery.[9-11] Early evidence has suggested reduced early failure rates through the adoption of revision knee networks.[12]

43 136

However, for some patients, this approach to managing patients is inevitably associated with increasing travel distances between patient's homes and their treating hospital. Travel distance has been shown to be an important factor in patient choice when selecting a surgeon for joint replacement surgery. It may be even more important for those awaiting revision joint replacement surgery as these patients struggle with mobility, may be unable to drive and may be more reliant on family members.[4] Evidence suggests that patients considering joint replacement are prepared to travel longer distances to obtain the best possible outcomes. A requisite in making such a decision requires data on outcomes of patients travelling greater

distances. Patients travel longer distances have been found to have higher readmission rates and higher mortality rates when undergoing other types of specialised surgery.[13] The pick-up rate of early complications, avoiding the need for readmission, may be less in areas further away from the main treatment centre. There is also concern that patients required to travel greater distances are more likely to be re-admitted to a different hospital than that where surgery was undertaken, resulting in clinical decisions that do not incorporate the primary surgeon and so potentially leading to poorer outcomes.[14] There is an absence of evidence in the literature to support or refute this argument in the context of patients undergoing RevKR. Therefore the aim of this paper is to investigate the relationship between longer patient travel distances and perioperative outcomes following RevKR performed in high volume tertiary referral centres.

159 Methods

161 Design

This study is a retrospective data analysis of observational data from the Hospital Episode Statistics (HES) and Office for National Statistics (ONS) databases. HES data is collected by NHS England for all patients treated at NHS hospitals in England and those treated at private hospitals where treatment was funded by the NHS. This study complies with the recommended reporting guidelines when using HES data[15] and the Strengthening of Reporting of Observational studies in Epidemiology (STROBE) guidelines.[16] Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

The analysis and presentation of data follows current NHS England guidance for the use of HES data for research purposes[17] and is anonymised to the level required by ISB1523 Anonymisation Standard for Publishing Health and Social Care Data.[18] The HES data were linked at a patient level to data from the ONS on deaths and date of death, which allowed the identification of patients who had died after their surgery. Linkage was achieved using a unique pseudonymised patient identifier using a previously validated methodology.[19] Patient travel distances were calculated using the Journey Time Statistics reference

Patient travel distances were calculated using the Journey Time Statistics reference
 document produced by the UK Department of Transport which modelled theoretical
 journey times between known centroids of Lower Layer Super Output Areas (LSOA)

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

of residence and NHS hospital sites.[20] Please refer to Supplementary material
S1 for Journey Times Statistics reference document.

Population An RevKR procedure was defined as a permanent removal or exchange of knee arthroplasty components. This includes a revision of a total knee replacement and a conversion of a unicondylar knee replacement to a total knee replacement. Secondary patellar resurfacing was not included as this represents a simple revision procedure, one that can be carried out in most nonspecialised hospitals. All patients aged \geq 18 years who underwent a RevKR in a high-volume trust between 1st January 2016 to 31st December 2019 were included in the study population. A high-volume trust was classified as a centre performing > 49 revisions per year. This revision volume threshold for classification represents that proposed by the British Association for Surgeons of the Knee (BASK) Revision Knee Working Group and is a mandatory requirement for all highly specialist centres co-ordinating regional networks. [21] As such centres attaining this threshold are more likely to represent tertiary referral centres where the stratification of more complex work will take place. Annual case volume at each trust was defined as the number of revision cases conducted in the year prior to the index procedure. This measure was preferred over a simple calculation of average annual volume as it accounts for recent experience at the point of surgery. The Office for Population Censuses and Surveys' Classification of Interventions and Procedures version 4 (OPCS-4) codes used to identify RevKR procedures are detailed in **Supplementary material S2**. Since laterality was needed to identify re-revisions, patients were excluded where the procedure laterality was not specified. The flow of patients, with numbers excluded at each point, is summarised in Supplementary material S3. To manage population heterogeneity, data were extracted for the period 1st April 2011 to 31st December 2019 and only the first revision for a specific side of the body record in this time period included. [22] Thus, any early revisions on the same side of the body in the four years and nine months preceding the start of the study period were identified and these patients excluded from the study. This aims to exclude the early revision knee replacement failures which have been shown to represent catastrophic failures potentially skewing our results.[22] We included revisions for infection as, despite

these representing a more variable patient group, presence of infection was thoughtto be unrelated to how far a patient lives from a specialised referral centre.

214 Exposure variable

Travel distances and times were calculated between a patient's LSOA and the postal codes for the treating hospitals. LSOAs are determined by the Office for National Statistics and are designed for the reporting of small area statistics. Public transport and highways data for England were used to create theoretical journey distances and times from origins to destinations. A network of journey distances and times from origins to destinations was produced using a software package called Transport Accessibility and Connectivity Calculator (TRACC). The Dijkstra's algorithm calculated the shortest route between these points. Data linkage was achieved with our clinical dataset following a reproducible workflow. The resulting travel distances and/or times for each patient were analysed as continuous variables. Three exposure variables were used. Straight line travel distance represented the distance "as the crow flies" between a patient's LSOA and treating hospital. Off peak driving distance represented the shortest driving distance between a patients LSOA and treating hospital. Finally peak driving times were calculated using average traffic speeds between 7am and 10am for the shortest possible road route between a patients LSOA and treating hospital. These three variables were used to account for variation in travel infrastructure between rural and urban areas and to attribute more meaningful results for patients.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

- ³ 235

Co-variates and cluster variable

The following groups of known or potential confounding variables were chosen apriori for inclusion in our multivariable logistic regression modelling:

Patient factors: Age in years (continuous), sex (male/female). Health co-morbidity was quantified using the Hospital Frailty Risk Score (HFRS). HFRS identifies frailty based on the occurrence of any of 109 International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10) codes used during any hospital admissions in the two years prior to, and for, the index admission. Deprivation was measured using the Index of Multiple Deprivation (IMD).[23] The

1

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

2 3	o 4 -	
4	245	IMD gives the LSOA where the patient lives a score based on a range of measures
5 6	246	of deprivation. IMD was analysed as a continuous variable.
7 8	247	Clinical factors: Defined by the presence or absence of infection as the primary
9 10	248	indication for RevKR. This was identified from the International Statistical
10	249	Classification of Diseases and Related Health Problems, tenth revision (ICD-10)
12 13	250	codes used during the admission.
14	251	Surgical factors: Surgeon and hospital volume (both continuous) was defined as the
16 17	252	number of RevKRs performed by a consultant or hospital in the 365 days prior to
18 19	253	each index procedure across the entire cohort. This was calculated before any
20 21	254	exclusion criteria was applied.
22 23	255	Temporal factors: Financial year of procedure (2015/16, 2016/17, 2017/18, 2018/19,
24 25	256	2019/20).
26 27	257	Hospital Provider: Clustering of patients by hospital provider was initially modelled
28 29	258	using random effects. However, despite variability between hospital providers with
30 21	259	primary and secondary outcomes, instability in the model estimates were observed.
32	260	To address the possibility of clustering at this level, a fixed effects model was
33 34	261	adopted with hospital provider as a covariate.
35 36	262	
37	263	
38 39	264	Outcomes
40	265	The primary outcome was emergency readmission within 30 days of discharge from
41 42	200	the index surgical beapital. Deadmission in this early period is your likely related to a
43	267	the index surgical hospital. Readmission in this early period is very likely related to a
44 45	268	complication of the surgical procedure. It has been used as a marker of perioperative
46 47	269	outcomes in similar studies investigating the relationship between patient travel
48 49	270	distance and outcomes following surgery. [13]
50 51	271	Secondary outcomes were:
52 53	272	90-day all-cause mortality, identified using linked data from Civil Registrations
54 55	273	(Mortality) dataset;
56 57	274	Inpatient length of hospital stay was attributed from continuous inpatient spells
58	275	(CIPS), which is the preferred estimate of length of stay. This refers to the length of
59 60	276	first stay after the operation regardless of any transfers across providers. The

60

BMJ Open

2		
3 4	277	median length of stay was calculated after visually inspecting the distribution and this
5 6	278	was dichotomized into prolonged length of stay if longer than the median stay.
7 0	279	
8 9 10	280 281	Statistical Analyses
11	282	Data was extracted from a secure, encrypted server controlled by NHS England.
12 13	283	Data were analysed within a secure, encrypted environment using standard
14 15	284	statistical software: R Studio version 2023.09.1+494 (Boston, Massachusetts, USA).
16 17	285	The R code and packages used are included in Supplementary material S4 .
18 19	286	Missing data were managed according to its extent and relevance to the aims of this
20 21	287	study. Age and IMD score were imputed for the small number of missing cases using
22	288	the mean of the entire study cohort. Given the central role of LSOA in estimating
23 24	289	travel distances and times and fewer than 5% of cases with missing data, these
25 26	290	cases were excluded to avoid the introduction of bias. Following data linkage,
27	291	approximately 36% (n = 5,838) of cases did not match with travel data. Multiple
28 29	292	imputation was performed using predictive mean matching based on the entire
30 31	293	cohort of patients with the following predictors: age, sex, HFRS score, IMD score,
32 33	294	hospital provider code, hospital volume and surgeon volume. Dependent variables
34	295	including readmission at 30 days, mortality at 90 days and length of stay were also
35 36	296	used in the imputation following a recommended approach using preditive mean
37 38	297	matching[24]. A total of five imputations were randomly chosen and subsequent
39 40	298	regression analyses were pooled.[25] Imputed data is shown in Supplementary
40 41 42	299	material S5.
43	300	Patient travel distances were categorised into quintiles for interpretation of baseline
45	301	demographics and clinical characteristics. Subsequent analysis of travel distances
46 47	302	and times were performed as continuous variables. Spearman's rank correlation was
48 49	303	performed to investigate the relationship between IMD score and patient age with
50 51	304	travel distances.
52 53	305	Straight line travel distance was modelled with restricted cubic splines to allow for

305 Straight line travel distance was modelled with restricted cubic splines to allow for
 306 the non-linear effects when testing the association with the primary outcome. All
 307 exposures were modelled with restricted cubic splines to allow for the non-linear
 308 effects when testing the association with prolonged length of stay. The Akaike

Information Criterion was used to select the most parsimonious specification of restricted cubic splines using the final adjusted model.

Fixed effects logistic regression models were used for the outcomes of readmission at 30 days, mortality at 90 days and prolonged length of stay. Adjustment for confounding was undertaken incrementally, adjusting for each of the five groups of confounding variables to explore their influence on the effect at each stage with reference to model fit statistics. This was done following an apriori methodology with addition and or removal of factors in the following order: patient factors, clinical factors, surgical factors, temporal factors and the hospital provider. The ultimate decision on the preferred statistical model was assessed using the Akaike Information Criterion (AIC) accepting the model with the lowest AIC. Co-variates were modelled as either linear or categorical terms to simplify the model and aid interpretability. Multicollinearity was assessed using eigenvalues, variance inflation factors and by examination of model parameter estimates with stepwise addition and removal of covariates. Odds ratios with 95% CIs and associated p-values were reported. A p-value of < 0.05 was taken to indicate statistical significance.

Results

Overview of results

A total of 16,736 patients met the inclusion criteria. Excluding missing LSOA data (n=171), 16,565 patients were included in the analysis. Following data linkage with department of transport journey times statistics, 10,727 patients had complete data linkage and data were imputed for the remaining 5,838 (35.2%). Of the 16,565 patients, 41.5% (n=6,880) presented to a tertiary referral centre and these data formed our analysis cohort. Patients were operated on across 181 hospital sites and 38 hospital trust providers. The baseline demographic and clinical characteristics of the patients were broadly similar between guintiles of straight-line travel distance. (Table 1). Higher hospital volumes were seen in patients travelling longer distances. Straight line travel distance was weakly correlated with age and social deprivation (Figure 1). Older patients were less likely to travel farther distances. Patients from the least deprived areas travelled shorter distances.

4.64

342						
43						
344	Table 1 – Bas	seline patient	demographi	cs and clinica	I characterist	ics stratified
45 46 47	by travel dista	ance quintiles	s from first im	puted datase	t	
947		Travel Distan	ce Quintile			
		1	2	3	4	5
	Distance	2.09 (1.35	4.42 (3.91	7.08 (6.34 to	11.39 (10.11	22.42 (18.09
	(Miles)	to 2.75)	to 5.00)	7.99)	to 12.74)	to 32.19)
	Driving Time	13 (9.3 to	20.45 (17 to	26.30 (21.98	34.10 (29.68	52.05 (42.68
	(Minutes)	17)	25)	to 31.13)	to 40.20)	to 66.83)
	Number of	1376	1376	1376	1376	1376
	patients					
	Tertiary	37 (97.37%)	38 (100%)	36 (94.74%)	35 (92.11%)	37 (97.37%)
	Providers					
	Age Mean	69.71	69.96	69.66 (10.92)	68.84 (11.01)	68.58 (10.75
	(SD)	(10.81)	(10.71)			
	Female Sex	762	768	729 (52.98%)	722 (52.47%)	734 (53.34%
		(55.38%)	(55.81%)			
	HFRS None	647	620	614 (44.62%)	666 (48.40%)	676 (49.13%
		(47.02%)	(45.06%)			
	HFRS Mild	438	474	485 (35.25%)	465 (33.79%)	433 (31.47%
		(31.83%)	(34.45%)			
	HFRS	241	236	243 (17.66%)	198 (14.39%)	230 (16.72%
	Moderate	(17.51%)	(17.15%)			
	HFRS Severe	50 (3.63%)	46 (3.34%)	34 (2.47%)	47 (3.42%)	37 (2.69%)

Page	14	of	1	16
------	----	----	---	----

1 2 3 4 5 6 7 8 9 10 11 22 3 4 5 6 7 8 9 10 11 21 31 4 15 16 17 8 9 20 21 22 32 4 25 26 27 8 9 30 31 32 33 4 35 6 37 38 37 37 38 37 38 37 38 37 38 37 38 37 38 37 38 37 38 37 38 37 38 37 37 37 37 37 37 37 37 37 37 37 37 37	
39 40	2.40
41 42 43	348 349 350
44 45	351
46 47	352
48 49	353 354
50 51	354 355
52 53 54	356 357
55 56	358
57 58	359
59 60	360

Infection	314	331	310 (22.53%)	334 (24.27%)	355 (25.80%)
Present	(22.82%)	(24.06%)			
Surgeon	7 (3 to 13)	7 (3 to 13)	8 (3 to 15)	8 (3 to 16)	9 (4 to 17)
Volume					
Hospital	73 (60 to	74 (60 to	79 (63 to 97)	79 (63 to 99)	85 (68.75 to
Volume	87)	89)			112)
IMD Score	16.44 (8.73	14.30 (7.96	14.50 (8.47	14.83 (9.23	14.752 (8.78
	to 28.67)	to 24.57)	to 21.36)	to 21.74)	to 21.45)
Year 2015/16	104 (7.56%)	94 (6.83%)	94 (6.83%)	89 (6.47%)	92 (6.69%)
Year 2016/17	383	354	348 (25.29%)	338 (24.56%)	353 (25.65%)
	(27.83%)	(25.73%)			
Year 2017/18	384	365	339 (24.64%)	360 (26.16%)	336 (24.42%)
	(27.91%)	(26.53%)			
Year 2018/19	269	325	347 (25.22%)	354 (25.73%)	339 (24.64%)
	(19.55%)	(23.62%)			
Year 2019/20	236	238	248 (18.02%)	235 (17.08%)	256 (18.60%)
	(17.15%)	(17.30%)			
			C	5.	
Outcomes					
The primary an	nd secondary	outcomes are	summarised in	n table 2.	
he observed i	rate of readmi	ssion at 30 da	ays was 8.3% (568/6880). Th	ere was a
egative assoc		in nigher strai	yni ine travel (er wide confid:	ance intervale	ennergency
etatical inferon	oo uayo (riyi	n bigher trave	al distance by	road and long	preciuueu ar drive timos

BMJ Open

2							
5 4	361	were not associated with statistically worse readmission rates at 30 days. The rate of					
5 6	362	mortality at 90 days v	vas only 3.2% (217/68	880). No statistically s	ignificant relationship		
$\frac{1}{7}$ 363 was observed between the distance a patient travels by road or the time a patient							
8 9	$\frac{3}{9}$ 364 spends travelling at peak driving times with rates of mortality at 90 days. 4						
10 11	365	(3421/6880) of patier	nts reported hospital s	tays more than 5 day	s. Following		
12	366	adjustment of confou	nding factors, we obs	erved no associations	s between prolonged		
13 14 15	367	length of stay and patient travel distance (Figures 3-5)					
16 17	368						
18	369	Table 2 – Adjusted	pooled Multivariable	e Logistic Regressio	n showing Odds		
19 20 21	370 371	Ratios for primary and secondary outcomes by exposure variables					
22 23	011		Straight line travel	Travel distance by	Peak Travel times		
24			distance (OR, 95%	shortest road route	by shortest road		
25 26 27			CI)	(OR, 95% CI)	route (OR, 95% CI)		
28 29		Readmission with	Figure 2	1.00 (0.99 to 1.02), p	1.00 (0.99 to 1.01), p		
30 31 32		30 days		value = 0.81	value = 0.69		
33 34		90 Day Mortality	1.00 (0.98 to 1.02), p	1.00 (0.99 to 1.01), p	1.00 (0.99 to 1.01), p		
34 35 36			value = 0.87	value = 0.86	value = 0.89		
37 38		Prolonged Length of	Figure 3	Figure 4	Figure 5		
39 40		stay					
41 42 43 44 45	372 373 374 375	•Odds ratios have been adjusted for patient age, sex, HFRS score,					
47 48 49 50	376 377 378	Discussion					
52 53	379	Statement of principal findings					
54 55	380	We present a multi-hospital site retrospective analysis of patients undergoing					
56	381	revision knee replace	ement surgery at tertia	ary referral centres in	England. In this		
57 58 59 60	382	analysis of 6,880 patients undergoing RevKR, we did not observe a statistical					
association between distance and time travelled for revision surgery and readmission within 30 days. Strengths and weaknesses of the study The findings of this study should be interpreted in view of several limitations. Firstly, this analysis used observational data from a large administrative dataset covering all NHS-funded procedures conducted in England. As with all administrative datasets we are limited in the amount of detail provided regarding presentation. We chose to categorise a high-volume centre by trust to accurately capture surgical experience. All NHS hospitals in England are run by hospital trusts which typically involve between one and four hospitals within a catchment area standardising their practice. It is common practice for specialist orthopaedic surgeons to move between these sites delivering the same procedures. Our study involved 187 hospital sites run by 38 trusts. We acknowledge this is a weakness of our study as this may not be representative of all trusts. We included all indications for RevKR in our patient cohort because indication was not thought to be related to how far a patient lives from a hospital. However, we acknowledge the rate of complications is higher in patients with infection and we subsequently adjusted for indication for revision in our analyses. [26] It is likely that because we did not exclude previous revision knee arthroplasty patients, the complexity of the surgery undertaken in our cohort varied. We recognise this is a limitation of the study however we assume case mix was unrelated patient travel distance. There were many missing patients (approximately 36%) following the linkage of HES data with Journey Time Statistics. To account for this, assumed that the data was

missing at random and used multiple imputation to estimate missing travel distances. It is likely the imputed values may introduce bias, however we modelled these based on predictors and dependent variables to improve our estimates. We do not present a sample size calculation, rather we have used all available data and our sample size was set by our inclusion criteria. We controlled for the clustered nature of our data between hospital providers through inclusion as a covariate in our modelling. To ensure consistency in our definition of tertiary referral hospitals, only hospitals performing >49 revisions/year were included. These are likely to treat a similar case

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Page 17 of 116

BMJ Open

416 mix of patients and potentially have similar access to resources within a national
417 healthcare system. This approach allowed us to control for variation across
418 providers. However, we acknowledge it does not fully account for the hierarchical
419 nature of the data with differences in treatment protocols and hospital specialisation
420 among factors which may influence patient outcomes.

There is a lack of granular data for revisions due to infection and therefore we excluded this patient group as some readmissions for this patient group may represent planned readmissions. There is also a lack of granular clinical data using HES for each readmission, therefore we cannot ascertain precise reasons for readmissions, but we assume are related to a post-surgical complication. Clinical coding practice within HES is known to vary across trusts.[27] As an example, some trusts may be more consistent in coding comorbidities, and this may have created some bias. However, this is unlikely to vary systematically with travel distances and so significantly bias our findings. We acknowledge the relatively short travel distances in this population compared to examples from the United States as such the results of this study may not be generalisable to larger geographical areas or less mature healthcare systems. However, the upper quintile in our study represents a substantial journey distance and time for our patient cohort where poor mobility is a significant factor affecting their care. This analysis does not consider journey times of those who may not have access to a car and instead chose to take public transport.

438 Strengths and weaknesses in relation to other studies, discussing important439 differences in results

⁴⁸41 This is the first study to analyse the potential impact of patient travel distances on
⁵⁰42 patients receiving RevKR. The findings that longer travel distances are not
⁵¹associated with inferior outcomes is an important part of the evaluation of the
⁵³assumptions and context behind the establishment of revision knee networks.[28]
⁵⁴This study has shown that concerns of introducing a network in larger geographical
⁵⁶regions, for example in Scotland where longer patient travel distances and times are
⁵⁸common, may be less important.[29] This is particularly useful as regions explore the

BMJ Open

2		
3 4 5 6 7 8 9 10 11 12 13 14 15	448	geography of their revision networks and during summative outcome assessment of
	449	this complex health intervention.[30] Despite there being a potential negative
	450	association between straight line travel distance and emergency readmission at 30
	451	days, there was a lack of association involving driving distances and times which
	452	present real world challenges for patients.
	453	It may be seen as surprising that no association between travel distance and
	454	prolonged length of hospital stay was identified. An expectation exists of increasing
16 17	455	difficulties being encountered with the discharge of patients living greater distances
18	456	from their treating hospital, which has been observed in patients following elective
19 20	457	pancreatic surgery.[31] This is also an observation seen in patients being treated in
21 22	458	specialist vascular centres in the United States which led to the recommendation of
22	459	additional care coordination and follow up efforts. However, the geography of the
24 25 26	460	population in these studies was much larger with significantly longer travel distances.
20 27 28	461	We did observe a weak but statistically significant correlation between social
28 29	462	deprivation status and age of the patient with longer travel distances. Patients from
30 31	463	poorer sociodemographic background may be expected to travel further for RevKR.
32 33	464	This highlights the additional care coordination and follow-up efforts that should
34	465	accompany the widening reach of regional revision knee networks. It is reassuring
35 36	466	that access to treatment for older patients is unaffected by travel distance. However,
37 38	467	there may be patients who refused to travel to a specialist centre and opted for
39 40	468	treatment at their local centre.
40 41 42 43	469	
	470	
44 45	471	
46 47	472	Meaning of the study: possible explanations and implications for clinicians and
48 49	473	policymakers
50 51	474	
52	475	The organisation and delivery of revision knee services in England has recently
53 54	476	undergone a substantial change and now such services are provided around
55 56	477	regional networks of care. This promises substantial advantages to the increasing
57 59	478	number of patients with problematic knee replacements in our ageing population who
59 60	479	will benefit from regional expertise.[8] However, it is unknown the impact of patients

Page 19 of 116

BMJ Open

1 2		
3	480	residing farther from tertiary referral centres, particularly rural patients who may
4 5 7 8 9 10	481	encounter additional difficulties associated with greater travel distance. A recent
	482	study following the outcomes of aortic surgery found that longer travel distances are
	483	associated with inferior perioperative outcomes[13]. Similar associations have been
	484	found in postoperative colorectal surgery patients [32]. As such our results are
12	485	reassuring to policy makers and clinicians.
13 14	10.6	
15 16 17	486	
17	487	Unanswered questions and future research
18 19	488 489	
20 21	490	There is a scarcity of evidence evaluating the patient perception of complex health
22	491	interventions such as network models of care. Recent work by Kugler et al has
23 24	492	demonstrated the willingness of patients to travel further for better outcomes in the
25 26	493	context of total knee replacement surgery. [33] Nevertheless, patient perceptions of
27 28	494	travelling further for their treatment should be a focus for future research in the
29	495	context of revision knee patients, particularly as this is one of the top ten research
30 31	496	priorities identified by the James Lind Alliance priority setting partnership.[34]
32 33	497	
34 35	400	
36 37	498	
38	499	Conclusion
39 40	500	
41 42	501	We did not observe an association in our study population between 30-day
43	502	readmission rates and increasing travel distances or times between a patient's home
44 45	503	and their treating hospital in revision knee replacement. This paper is the first to
46 47	504	explore the relationship between travel distance and complex orthopaedic surgery
48 ⊿q	505	and informs some concerns regarding the creation of a centralised revision knee
50	506	network. This information is of utility to surgical providers and commissioners of
51 52	507	healthcare services. Furthermore, it can inform patient-led decision making and the
53 54	508	exploration of perceptions surrounding travelling for complex surgery. Although this
55	509	is the first assessment in complex orthopaedic surgery, a prospective analysis will be
50 57	510	undertaken as part of the ongoing auditing of revision knee networks in England.
58 59 60	511 512	
	U 1 H	

2 3		
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17	513 514 515	Supplementary material and figures
	516 517 518	Supplementary material S1 – Journey Time Statistics Reference Document
	519	Supplementary material S2 – OPCS-4 code criteria used for Hospital Episode
	520 521 522	Statistics data extraction
18 19	523	See separate file named supplementary material S2
20 21 22 23	524 525	
24 25	526	
26 27	527	Supplementary material S3 – Flow of patient inclusion/exclusions
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42	528 529 530	-See attached file named Supplementary Material S3
	531	
	532	Supplementary material S4 – R Code
	533 534	See attached file named Supplementary Material S4
	535 536	
	537	Supplementary material S5 –Scatterplot for imputed data: A comparison
43 44	538	between imputed values and observed values following multiple random
45 46	539	imputation. Imputed values in "blue", observed values in "grey". Imputation 0
47 48	540	on X axis refers to original dataset. Subsequent random imputations labelled 1
49 50 51 52 53 54 55 56 57 58 59 60	541 542 543 544 545 546 547	to 5 on x axis.

BMJ Open

2 3		
4	548	Figure 1 -
5 6 7 8	549	
	550	(Left) Scatterplot showing correlation between patient age and travel distance.
9 10	551	Red line represents linear regression trend. Spearman's rank correlation is
11 12	552	presented in chart.
13 14	553	
15 16	554	(Right) Scatterplot showing correlation between social deprivation and patient
17 18	555	travel distance. Red line represents linear regression trend. Spearman's rank
$\begin{array}{c} 18\\ 19\\ 20\\ 21\\ 22\\ 32\\ 4\\ 25\\ 26\\ 27\\ 28\\ 29\\ 30\\ 31\\ 32\\ 33\\ 34\\ 35\\ 36\\ 37\\ 38\\ 90\\ 41\\ 42\\ 43\\ 44\\ 56\\ 47\\ 48\\ 9\\ 50\\ 51\\ 52\\ 53\\ 56\\ 57\\ 58\\ 90\\ \end{array}$	556 557 558 559 560 561 562 563 564 565 566 567 568 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586	correlation is presented in chart.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

1

2		
3 4	587	Figure 2 - Predicted probability of emergency readmission at 30 days by
5 6 7 8	588	straight line patient travel distance from hospital after RevKR
	589	A Fixed effects multivariable logistic regression model using 3 knots at 5%,
9 10	590	50% and 95% centiles of mean unit volume. 95% confidence intervals
11 12 13 14	591 592 593	represented by blue shaded line
15 16	594	
17 18	595	
19 20 21	596	
22 23	597	
24 25	598	
26 27	599	Figure 3 - Predicted probability of prolonged length of inpatient stay at by
28 29	600	patient straight line travel distance from hospital after RevKR
30 31	601	A Fixed effects multivariable logistic regression model using 4 knots at 5%,
32	602	35%, 65% and 95% centiles of mean unit volume. 95% confidence intervals
33 34	603	represented by blue shaded line
35 36	604	
37 38	605	
39 40	606 607	
41 42	608	Figure 4 - Predicted probability of prolonged length of inpatient stay at by
43 44	609	patient driving distance from hospital after RevKR
45 46	610	A Fixed effects multivariable logistic regression model using 4 knots at 5%,
47 48	611	35%, 65% and 95% centiles of mean unit volume. 95% confidence intervals
49 50	612	represented by blue shaded line
51 52	613	
53 54	614 615	
55 56		
57 58		
59 60		
00		

1 2		
3 4	616	Figure 5 - Predicted probability of prolonged length of inpatient stay at by
5 6	617	patient driving time from hospital after RevKR
7 8	618	A Fixed effects multivariable logistic regression model using 4 knots at 5%,
9 10	619	35%, 65% and 95% centiles of mean unit volume. 95% confidence intervals
11 12	620	represented by blue shaded line
13 14	621	
15 16	622	
17		
10 19		
20 21		
22		
24 25		
26 27		
28 29		
30 31		
32 33		
34 35		
36 37		
38 39		
40 41		
42 43		
44 45		
46 47		
48 49		
50 51		
52 53		
54 55		
56 57		
58 59		
60		

1 2		
3	623	
5	624	
6 7	625	
8	626	
9 10 11	627	Contributorship
12 13	628	
14 15	629	
16 17	630	Alex Matthews: Conceptualisation, Methodology, Project Administration,
17	631	Investigation, Data Curation, Formal Analysis, Visualisation, Writing - original draft,
19 20	632	Writing - review and editing. This author is the guarantor and is responsible for the
21 22	633	content
23	634	
24 25	635	Jonathan P Evans: Conceptualisation, Supervision, Writing - review & editing
26 27	636	
28	637	Jonathan T Evans: Supervision, Writing - review and editing
29 30	638	
31 32	639	Sarah E Lamb: Conceptualisation, Supervision, Writing - review and editing
33 34	640	
35	641	Andrew Price: Conceptualisation, Supervision, Writing - review and editing
36 37	642	
38 39	643	William Gray: Conceptualisation, Supervision, Methodology, Writing - review and
40	644	editing
41 42	645	
43 44	646	Tim Briggs: Supervision, Writing - review and editing
45 46	647	
47 48	648	Andrew Toms: Conceptualisation, Supervision, Writing - review and editing
49 50 51	649	
52 53 54	650	
55	651 652	
56 57	653	
58	654	
59 60	655	
	030	

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1 2		
3 4	657	F
5	658	
7	659	٦
8 9	660	F
10 11	661	C
12	662	r
13 14	663	1
15 16	664	C
17	665	t
18 19 20 21 22	666	
23 24	667	(
25	668	
26 27	669	7
28 29	670	ç
30	671	e
31 32	672	â
33 34	673	â
35 36 37	674	
38 39	675	(
40	676	
41	670 677	A
43 44	678	r
45 46	679	C
47	680	t
40 49	681	C
50 51	682	V
52 52	(02	
55 54	687 687	
55 56	004	
57	685	
58 50	686	
- 60	000	

Public and Patient Involvement statement

The study's chief investigator (AT) led the James Lind Alliance 'Revision Knee Replacement' priority setting partnership. This group of patients, carers and health care professionals identified the need to investigate the best way of organising revision knee replacement surgery to improve patient outcomes as one of their top 10 research questions. Patients were therefore directly involved in the development of the study's aims and objectives. The results of the study will be disseminated to the members of this group prior to publication.

7 Copyright/licence for publication Statement

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence.

675 Competing Interests

All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

685 Transparency Declaration

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

The lead author (the manuscript's guarantor) affirms that the manuscript is an

honest, accurate, and transparent account of the study being reported; that no

important aspects of the study have been omitted; and that any discrepancies from

the study as planned (and, if relevant, registered) have been explained.

Ethical Approval

The analysis and presentation of data follows current NHS England guidance for the use of HES data for research purposes and is anonymised to the level required by ISB1523 Anonymisation Standard for Publishing Health and Social Care Data. Ethical approval was not required.

Funding

No funding was obtained to carry out this study

Data Sharing

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

2					
3	718	1. Price AJ, Alvand A, Troelsen A, Katz JN, Hooper G, Gray A, et al. Knee replacement.			
4	719	Lancet, 2018:392(10158):1672-82.			
5	720	2 Sabah SA Hedge FA Abram SGE Alvand A Price AL Honewell S Patient-reported			
7	721	outcome measures following revision knee replacement: a review of PROM instrument			
8	721	utilization and measurement properties using the COSMIN shead list. DNU Oper			
9	722				
10	725	2021,11(10).6040109.			
11	724	3. Saban S, Knight R, Alvand A, Murray D, Petrou S, Beard D, Price A. No exponential			
12	125	rise in revision knee replacement surgery over the past 15 years: an analysis from the			
13	726	National Joint Registry. Osteoarthritis and Cartilage. 2022;30(12):1670-9.			
14	727	4. Rowland T, Lindberg-Larsen M, Santy-Tomlinson J, Jensen CM. A qualitative study of			
16	728	patients' experiences before, during and after surgical treatment for periprosthetic knee			
17	729	joint infection; "I assumed it had to be like that ". International Journal of Orthopaedic			
18	730	and Trauma Nursing. 2023;48:100992.			
19	731	5. Garbarino LJ, Gold PA, Sodhi N, Anis HK, Ehiorobo JO, Boraiah S, et al. The effect of			
20	732	operative time on in-hospital length of stay in revision total knee arthroplasty. Ann Transl			
21	733	Med. 2019:7(4):66.			
22	734	6 Liodakis E Bergeron SG 7ukor DL Huk OL Enure LM Antoniou L Perioperative			
24	735	Complications and Length of Stay After Revision Total Hin and Knee Arthroplasties: An			
25	736	Analysis of the NSOID Database. The Journal of Arthroplasty, 2015;20(11):1868, 71			
26	730	7 Driggs TWD A National Daviaw of A dult Elective Orthonactic Services in England			
27	121	7. Briggs I WK. A National Review of Adult Elective Orthopaedic Services in England.			
28	/38	In: Time GIRF, editor. 2015.			
29	/39	8. Kaison NS, Matnews JA, Phillips JRA, Baker PN, Price AJ, Toms AD, et al. Revision			
30 31	/40	knee replacement surgery in the NHS: A BASK surgical practice guideline. The Knee.			
32	/41	2021;29:353-64.			
33	742	9. Yapp LZ, Walmsley PJ, Moran M, Clarke JV, Simpson AHRW, Scott CEH. The effect of			
34	743	hospital case volume on re-revision following revision total knee arthroplasty. J Bone Joint			
35	744	Surg Br. 2021;103-B(4):602-9.			
36	745	10. Halder AM, Gehrke T, Günster C, Heller KD, Leicht H, Malzahn J, et al. Low Hospital			
3/ 20	746	Volume Increases Re-Revision Rate Following Aseptic Revision Total Knee Arthroplasty: An			
30 39	747	Analysis of 23,644 Cases. J Arthroplasty. 2020;35(4):1054-9.			
40	748	11. Jeschke E, Gehrke T, Günster C, Heller KD, Leicht H, Malzahn J, et al. Low Hospital			
41	749	Volume Increases Revision Rate and Mortality Following Revision Total Hip Arthroplasty: An			
42	750	Analysis of 17.773 Cases, J Arthroplasty, 2019;34(9):2045-50.			
43	751	12. Bloch BV, Matar HE, Berber R, Grav WK, Briggs TWR, James PL, Manktelow ARJ, The			
44 45	752	impact of a revision arthroplasty network on patient outcomes. I Bone Joint Surg Br			
45 46	753	2023·105-b(6)·6/1-8			
47	757	12 Eoldman 7M Zhong V Mao I Sumpio BI Mohohali I Chang DC et al Longer nationt			
48	754	15. Feldinan Zivi, Zheng X, Wao J, Sumpio BJ, Woneban J, Chang DC, et al. Longer patient			
49	155	traver distance is associated with increased non-index readinission after complex aortic			
50	/56	surgery. Journal of Vascular Surgery. 2023;77(6):1607-17.e7.			
51	/5/	14. Saunders RS, Fernandes-Taylor S, Kind AJ, Engelbert TL, Greenberg CC, Smith MA, et			
52 52	/58	al. Rehospitalization to primary versus different facilities following abdominal aortic			
55 54	759	aneurysm repair. J Vasc Surg. 2014;59(6):1502-10, 10.e1-2.			
55	760	15. Sinha S, Peach G, Poloniecki JD, Thompson MM, Holt PJ. Studies using English			
56	761	administrative data (Hospital Episode Statistics) to assess health-care outcomessystematic			
57	762	review and recommendations for reporting. Eur J Public Health. 2013;23(1):86-92.			
58					
59 60					
00					

BMJ Open

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

16. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344-9. 17. NHS Digital. Hospital Episode Statistics(HES) Analysis Guide. London, UK: NHS Digital; 2018. 18. Information Standard Board for Health and Social Care. Anonymisation Standard for Publishing Health and Social Care Data Specification. London, UK: NHS Digital; 2013. NHS Digital. Delivering linked datasets to support health and care delivery and 19. research. London, UK: NHS Digital; 2018. 20. Department for Transport. Journey Time Statistics, notes and definitions. London, UK: Department for Transport; 2019. Kalson NS, Mathews JA, Miles J, Bloch BV, Price AJ, Phillips JRA, et al. Provision of 21. revision knee surgery and calculation of the effect of a network service reconfiguration: An analysis from the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. Knee. 2020;27(5):1593-600. 22. Deere K, Whitehouse MR, Kunutsor SK, Sayers A, Price AJ, Mason J, Blom AW. How long do revised and multiply revised knee replacements last? An analysis of the National Joint Registry. Lancet Rheumatol. 2021;3(6):e438-e46. 23. Kamath CC, O'Byrne TJ, Lewallen DG, Berry DJ, Maradit Kremers H. Neighborhood-Level Socioeconomic Deprivation, Rurality, and Long-Term Outcomes of Patients Undergoing Total Joint Arthroplasty: Analysis from a Large, Tertiary Care Hospital. Mayo Clin Proc Innov Qual Outcomes. 2022;6(4):337-46. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple 24. imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Medical Research Methodology. 2017;17(1):162. 25. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Ann Transl Med. 2016;4(2):30. 26. Sabah SA, Knight R, Alvand A, Palmer AJR, Middleton R, Abram SGF, et al. Patient-Relevant Outcomes Following First Revision Total Knee Arthroplasty, by Diagnosis: An Analysis of Implant Survivorship, Mortality, Serious Medical Complications, and Patient-Reported Outcome Measures Utilizing the National Joint Registry Data Set. J Bone Joint Surg Am. 2023;105(20):1611-21. 27. Hardy F, Heyl J, Tucker K, Hopper A, Marchã MJ, Briggs TWR, et al. Data consistency in the English Hospital Episodes Statistics database. BMJ Health Care Inform. 2022;29(1). 28. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021;374:n2061. Farrow L, Kennedy JW, Yapp L, Harding T, Walmsley P. Provision of revision knee 29. arthroplasty services across Scotland: A national audit. Knee. 2023;42:312-9. Paterson C, Baarts C, Launsø L, Verhoef MJ. Evaluating complex health interventions: 30. a critical analysis of the 'outcomes' concept. BMC Complement Altern Med. 2009;9:18. Jackson KL, Glasgow RE, Mone MC, Sheng X, Mulvihill SJ, Scaife CL. Does travel 31. distance influence length of stay in elective pancreatic surgery? HPB (Oxford). 2014;16(6):543-9. 32. Cloyd JM, Huang L, Ma Y, Rhoads KF. Predictors of readmission to non-index hospitals after colorectal surgery. The American Journal of Surgery. 2017;213(1):18-23.

1 2		
3 4 5 6 7 8 9 10 11 12	809 810 811 812 813 814 815	 33. Kugler CM, De Santis KK, Rombey T, Goossen K, Breuing J, Könsgen N, et al. Perspective of potential patients on the hospital volume-outcome relationship and the minimum volume threshold for total knee arthroplasty: a qualitative focus group and interview study. BMC Health Serv Res. 2021;21(1):633. 34. Mathews JA, Kalson NS, Tarrant PM, Toms AD. Top ten research priorities for problematic knee arthroplasty. Bone Joint J. 2020;102-b(9):1176-82.
13 14 15 16 17 18 19		
20 21 22 23 24 25 26		
20 27 28 29 30 31 32		
33 34 35 36 37 38		
39 40 41 42 43 44 45		
46 47 48 49 50 51		
52 53 54 55 56 57		
58 59 60		

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open







30 35 40 45 Travel Distance (Miles)

Figure 2

87x59mm (300 x 300 DPI)

50

55

60

65

70

75

20

25

30



BMJ Open: first published as 10.1136/bmjopen-2024-085201 on 6 May 2025. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

BMJ Open



Figure 3 76x59mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



Figure 4

BMJ Open: first published as 10.1136/bmjopen-2024-085201 on 6 May 2025. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

78x60mm (300 x 300 DPI)



BMJ Open







72x60mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Journey Time Statistics:

Notes and Definitions

About this

release

This publication supports the latest statistics on journey times.

In this publication

Overviewp1
Access to key services
p4
Connectivityp7
Data sourcesp9
Outputsp18
Strengths and weaknesses
p19

Further information

Public enquiries 020 7944 3077 vehicles.stats@dft.gov.uk Media enquiries 020 7944 3066

Overview

This note provides information on the methodology used, the source data and definitions of key terms for calculating Journey Time Statistics.

These annual statistics were first published in December 2015 for the year 2014 and have been developed from the earlier Accessibility Statistics published for 2007 to 2013.

The Journey Time Statistics produced by DfT consists of theoretical journey times calculated by modelling journeys between known sets of origins and destinations. It uses information on the road network, traffic speeds and public transport timetables in England.

The relevant Journey Time Statistics calculation is varied for origins and destination to meet a variety of needs. Two sets of analysis are published: Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

- Access to key services; and
- Connectivity

Origin indicators

These indicators measure the number of different services in a particular area that users can reach within a given time.

Destination indicators

These indicators measure the proportion of users that can access a service within a certain time.

The 'user' populations for each service in the destination indicators are:

Employment	16-74 year olds
Primary schools	5-10 year olds
Secondary schools	11-15 year olds
Further education	16-19 year olds
All other services	All households

Journey Time Statistics: Notes and Definitions - Page 1

Key services

- Employment centres: Data used are the number of jobs in a Lower Super Output Area (LSOA). The data tables include results for employment centres of 3 different sizes (100-499 jobs, 500-4,999 jobs and at least 5,000 jobs). For the key services average, the 500-4,999 jobs definition is used for employment.
- <u>Education</u>: Locations of all open Primary schools, Secondary schools, Further Education and Sixth Form Colleges.
- General Practice (GP) surgeries: For 2017 based on the Patients Registered at a GP Practice dataset released by NHS Digital – previously this was based on a filtered dataset of NHS prescribers released by NHS Digital.
- Hospitals: Based on hospitals that are registered with the Care Quality Commission (CQC) and are managed by Acute Trusts.
- ► <u>Food stores</u>: Locations of grocery, supermarkets or convenience stores.
- Town centres: Locations of Town centres using a central focal point for the town mapped to the nearest road.

Geography

Local authorities

In some parts of England there are two tiers of local authorities, and in others a single unitary authority. Statistics have been calculated for both types of authority - around 360 in all. These vary considerably in size, from a population of a few tens of thousands to over a million.

Lower Layer Super Output Areas (LSOA)

LSOAs are small areas designed to be of a similar population size, with an average of approximately 1,500 residents or 650 households. There are 32,844 Lower-layer Super Output Areas (LSOAs) in England. They were determined by the Office for National Statistics for the reporting of small area statistics and are derived from the 2011 Census.

Urban and rural definitions

This report uses the Defra Rural-Urban Classification, based on 2011 Census Output Areas. The Rural-Urban Classification defines areas as rural if they fall outside of settlements with more than 10,000 resident population. See <u>Defra's Definitions and Local Authority Classification</u> for more details.

Journey time calculations

The journey time calculations are carried out using a commercially available software package 2 3 called TRACC, owned by Basemap. TRACC is a desktop application that uses public transport and 4 highways data to create journey times from origins to destinations. It uses timetable information 5 6 showing both arrival and departure times at stops from public transport services against a specific 7 time/day period. Highways information from road networks are used to fill the gaps between public 8 9 transport services by creating a linear network that connects the origins, destinations and stops 10 together. This provides a fully routable network of nodes and lines which is saved on file as a graph 11 12 network. The graph network has various constraints which can be altered to suit the user need 13 such as distance travelled, interchange delays on public transport and stopping limitations on road 14 15 networks. The TRACC software then gueries the graph network with origin and destination co-16 17 ordinates and uses the Dijkstra shortest path algorithm to route between these points. This is an 18 algorithm for finding the shortest distance for travel between the graph networks. 19

20 For a public transport journey, the journey time produced includes all walking elements of the 21 journey, i.e. the walk from the origin of the journey to the road, from the road to public transport 22 23 stops, any interchange of public transport using the road and then from the final stop to the 24 destination via the road, and finally from the nearest point on the road network to the destination. 25 26 The journey assumes arrival at the first stop one minute before the initial departure, with any 27 subsequent interchange waiting times included as part of the final journey time. 28 29

Car, cycle or walk only journeys are similar except that once the road network is reached the
 journey proceeds link by link along the road network at speeds governed by data held in the model.
 These are specific to the mode, the road type, and in some cases the individual road link.

The 10 shortest journey times from each origin (i.e. Output Area) are calculated for each destination type. For the public transport / walking mode these consist of the 10 shortest journey times by either walking or public transport, after applying a 5 minute penalty for any journeys using public transport (to represent travellers arriving slightly early at the first stop).

The journey times are representative of the 'morning peak'. This is made explicit for public transport / walking by requiring the journey to be completed between 7 and 10am, and for car journeys by using average traffic speeds for between 7 and 10am. For the cycle mode no actual speed data are available. The cycle speeds used are default assumptions, and are not based on a particular time of day.

Access to key services

The Access to Services analysis applies the Journey Times methodology to origins consisting of residential neighbourhoods and destinations consisting of centres of employment and a range of key local services. Journey times are calculated for three modes of transport: public transport; driving; and cycling. These journey times are then used to generate further indicators, as described in the Outputs Section.

The Access to Services calculation process and the coverage of the data set are very similar to those of the Accessibility Statistics from which they were developed. However, the calculation algorithm and a number of other features of the design are different, so the results are not directly comparable.

The statistics are designed to represent as much as possible the situation on a Tuesday in October of the year to which they relate. Data for the second week of October are used in the analysis, since this provides a fairly typical week, unaffected by major national holidays, school holidays or other seasonal effects. The origins, destinations and public transport timetables used are as far as possible for this date. The traffic data are averages for the preceding 12 months up to and including August. The road networks are those current at the start of the traffic data year.

Outline of access to services calculation process

Origins 171,372 Output Areas (OA) (Census geography)

Destinations Employment locations (3 sizes) Education (Primary schools, Secondary Schools, Further Education colleges) Health (GPs, Hospitals) Food stores Town centres

> **Transport data Bus/rail timetables** Road network Average road speeds

Travel time calculation Using TRACC software, similar to running millions of journey planner queries



Output data

Travel times from each of 32,844 Lower Super Output Areas (LSOA) to nearest 10 of each destination

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Public transport / walk x1 time period AM peak

x3 modes Cycle Car

Model parameters and assumptions

General parameters

Maximum journey time of 2 hours.

Maximum journey distance of 100km.

Walking

60

These apply to both:

walking between origin / destination and the transport networks at both ends of a journey by

Journey Time Statistics: Notes and Definitions - Page 4

2 3

4

5 6

7

8

any mode;

walk only journeys as part of the public transport / walk mode.

Maximum straight line distance between origin / destination and road network of **2km**. The algorithm will always use nearest point on network. For cycle or car modes, travel by cycle or car begins from this point. For public transport/walk, traveller walks along road network to the most suitable public transport stop, or direct to the destination if this is quicker.

- ⁹ Walking speed on road/path network of **4.8km/h.**
- Walking speed off road/path network of **4.0km/h**.

13 **Public transport**

Interval within which door-to-door journey must be completed (required for timetable selection) is
 7am to 10am on a Tuesday.

¹⁷Maximum walk distance of **3km** - this applies to walks from origin to first public transport stop, from ¹⁹last stop to destination, and also walking directly from origin to destination without using public ²⁰transport at all.

Maximum number of potential first public transport stops considered in routing algorithm is 100 (starting with the closest to origin).

Allowance for catching first public transport service is **5 minutes** - added to any journey that involves boarding one or more public transport services.

Public transport speed – this is provided implicitly by the timetable information.

Interchange time of 5 minutes (minimum interval allowed between arriving at a stop and catching
 another service).

 $^{33}_{34}$ Maximum straight line distance between public transport interchanges of **500m**.

Stop clustering at 150m – groups together public transport stops within this distance of one another
 to speed up processing. The individual timetables for each service are retained.

Cycling speeds

38 39 40

58 59 60

Road Type	Speed
Motorway	0.0 km/h
Urban Motorway	0.0 km/h
A road	16.0 km/h
B road	16.0 km/h
Minor road	16.0 km/h
Local street	16.0 km/h
Private road – restricted access	4.8 km/h
Private road – public access	16.0 km/h
Pedestrian street	4.8 km/h
Alley	4.8 km/h

Parking time of **5 minutes** - added to all cycle journeys.

Car speeds

-				
Type of road	2014	2015	2016	2017
		Default spe	eeds (km/h)	
Motorway	79.5	77.0	77.5	77.6
Urban Motorway	79.5	77.0	77.5	77.6
A road	42.7	43.7	43.3	43.2
B road	41.6	43.0	42.2	41.9
Minor road	36.8	37.5	36.8	36.3
Local street	19.2	17.8	18.8	18.3
Private road – restricted access	17.0	16.7	16.2	15.3
Private road – public access	14.8	15.2	15.1	13.6
Pedestrian street	0.0	0.0	0.0	0.0
Alley	0.0	0.0	0.0	0.0

Car speeds are calculated for specific links where more than 200 records exist otherwise the default speeds are used. Minimum journey time for a journey that uses a car is **5 minutes**.

Time at junctions

Road normalisation is used for all modes of transport which converts each road link to a straight line to speed up processing. The true link length is retained for accurate speed/time calculations, but there could be a small effect on the calculation of shortest distance from the road network to destination points. Effect for origins is minimal due to origins being constrained to road nodes.

Connectivity

These experimental analyses are intended to apply the Journey Times methodology to a range of more strategic or economically significant destinations than the primarily local services covered by the Access to Services analyses; including airports and railway stations. The principle difference in the Connectivity approach from that of the Access to Services analyses is that journey times are calculated, as far as possible, to all accessible locations, rather than to just the nearest 10 examples. This tends to result in a much larger data set being generated. In some cases a longer maximum journey time may be allowed although this may depend on what is considered reasonable for the type of destination. Given these factors, a less detailed origin data set may be used than for Access to Services. This is both necessary, to limit the size of the data set, and acceptable where the typical journey lengths are longer.

The first connectivity analyses published using the new Journey Time methods were released in Journey Time Statistics 2015, published in April 2017, for two destination sets – airports and rail stations. These analyses using the Journey Times methods superseded two earlier Connectivity Statistics reports published in 2014 and 2015 based on the old accessibility statistics methods, in the same way that the new Access to Services analyses have replaced the earlier Accessibility Statistics. Again, the connectivity results produced using the old and new methods are not directly comparable.

Outline of Connectivity calculation



Model parameters and assumptions

Origins	Population weighted centroids (the central
	point) of 32,844 English LSOAs as specified in
	the 2011 Census geography. These points were
	then constrained to the nearest road node, as
	for Access to Services method.

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

BV	AJ Open Page
Journey Time Calculation	As for Access to Services, for public transport
	/ walking and car modes only, except that a
	maximum journey time of 240 minutes and
	maximum straight line distance of 400km is
	allowed.
Outputs	Generally similar to Access to Services,
•	with different journey time classifications as
	appropriate Journey time results to specific
	destinations are included – this is the key
	difference in the Connectivity analyses
	(Average inverse) times' and (nearest)
	Average journey times and nearest
	destinations should be used with caution.
	The average journey times exclude results
	for areas with no available connection under
	240 minutes, which may become significant
	in remote areas and for destinations are a
	great distance from the origin. The 'nearest'
	destination is the destination with the shortest
	average journey time across the whole area
	appointed which will be relatively large in the
	sees of least with arith least use the
	case of local authority level results.

review only

http://bmjopen.bmj.com/site/about/guidelii

Data sources

Origins

The origins used for all Access to Services calculations are the 171,372 English Output Areas (OA) as specified in the 2011 Census geography.

To provide the actual journey start point in each OA, the population weighted centroid of the OA was shifted to the nearest node (i.e. junction) on the road network. This was to avoid biasing the journey time results where the centroid of the OA was a long way from a road. In fact it is rare for an OA centroid to be more than about 100 metres from a road – only a tiny handful of OA in remote areas have centroids as much as 1km from a road. The OA centroids have been shifted onto the nearest road node rather than the nearest point on a road in order to reduce issues arising from normalising the road network.

Origin	Data source for the origin points
All	Data: Population centroid of each Output Area in
	2011.
	Source: ONS 2011 Census Boundaries.
	Further information: <u>http://geoportal.statistics.gov.ul</u>

Destinations

The destinations used consist of three different sizes of employment centre and the locations of seven other types of key local service. For each of these key services a nationally consistent data set has been identified or derived – further information on these is provided in this section.

Each destination is located by a 6-figure National Grid reference. For the employment destinations this is taken to be the population weighted centroid of the LSOA.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

Destination	stination Number of locations			
	2014	2015	2016	2017
Employment centres (small)	16,465	16,625	16,930	17,194
Employment centres (medium)	9,235	9,460	9,707	10,241
Employment centres (large)	645	676	719	785
Primary schools	16,463	16,484	16,655	16,927
Secondary schools	3,365	3,376	3,381	3,174
Further education colleges	2,624	2,606	2,418	2,304
GPs	9,257	11,167	9,128	7,353
Hospitals	296	278	278	277
Food stores	19,549	19,746	21,665	20,987
Town centres	1,211	1,211	1,211	1,211

The data source for GP surgeries was reviewed and replaced for 2017.

Access to key services

Destinations 2017	Data source for the locations	Data source for users of the
	of the service	service
Employment	Data: Number of jobs available in a LSOA in the year before the calculation year.	Data: Number of 16-74 year olds in each output area.
	Source: ONS Business Register Employment Survey.	Source: ONS mid-year population estimates for calculation year.
	Further information: <u>https://</u> www.nomisweb.co.uk/default. asp	Further information: ONS mid-year population estimates: http://www.ons.gov.uk/ons/ taxonomy/index.
Primary schools	Data: Location of all open primary schools in September of calculation year.	Data: Number of 5-10 year old in each output area.
	Source: The Department for Education (DfE) Edubase.	Source: ONS mid-year population estimates for calculation year.
	Further information: <u>https://get-information-schools.service.</u> gov.uk/	Further information: ONS mid-year population estimates: http://www.ons.gov.uk/ons/ taxonomy/index.
Secondary schools	Data: Location of all open secondary schools in September of calculation year.	Data: Number of 11-15 year olds in schools in each output area.
	Source: DfE Edubase.	Source: ONS mid-year population estimates for calculation year.
	Further information: <u>https://get-information-schools.service.</u> <u>gov.uk/</u>	Further information: ONS mid-year population estimates: http://www.ons.gov.uk/ons/ taxonomy/index.
Further education colleges	Data: Location of all open further education and sixth form colleges/school sixth form in September of calculation year.	Data: Number of 16-19 year olds in each output area.
	Source: DfE Edubase.	Source: ONS mid-year population estimates for calculation year.
	Further information: <u>https://</u> get-information-schools. service.gov.uk/	Further information: ONS mid-year population estimates: http://www.ons.gov.uk/ons/ taxonomy/index.

Journey Time Statistics: Notes and Definitions - Page 10

For peer review only - http://bmjopen.bmj.com/site/about/guidelin

Destinations 2017	Data source for the locations	Data source for users of th
	of the service	service
GPs	Data: Locations of GP	Data: Number of households
	surgeries with registered	each output area.
	patients in October of	
	calculation year	
	Source: NHS Digital table of	Source: 2011 Census +
	Registered natients at GP	Local Authority (LA) undates
	practices	from the Ministry of Housing
	practices	Communities & Local
		Covernment (MHCLC) mid
		Government (MinCLG) Ind-
	Further information: <u>https://</u>	Further information: 2011
	digital.nns.uk/data-and-	Census: <u>http://www.nomiswe</u>
	Information/publications/	<u>co.uk/census/2011</u>
	statistical/patients-registered-	MHCLG mid-year household
	at-a-gp-practice	projections: https://www.gov
		uk/government/statistical-da
		sets/live-tables-on-househol
		projections
Hospitals	Data: Location of hospitals.	Data: Number of households
		each output area.
	Source: Care Quality	Source: 2011 Census + LA
	Commission - Directory of	updates from MHCLG mid-
	places that provide care.	year household projections of
		calculation year.
	Further information: http://www.	Further information: 2011
	cac.org.uk/content/how-get-	Census: http://www.nomiswe
	and-re-use-coc-information-	co.uk/census/2011
	and-data	
		MHCLG mid-year nousenoid
		projections: <u>nttps://www.gov</u>
		<u>uк/government/statistical-da</u>
		sets/live-tables-on-househol
		projections

http://bmjopen.bmj.com/site/about/guid

review only

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Destinations 2017	Data source for the locations	Data source for users of th
	of the service	service
Food stores	Data: Location of grocery/	Data: Number of households
	supermarkets or convenience	each output area.
	stores in October of calculation	
	year.	
	Source: The Local Data	Source: 2011 Census + LA
	Company	updates from MHCLG mid-
		vear household projections of
		calculation year
	Further information: https://	Further information: 2011
	www.localdatacompany.com/	Census: http://www.nomiswe
		co.uk/consus/2011
		MHCLG mid-year household
		projections: <u>https://www.gov.</u>
	4	uk/government/statistical-dat
		sets/live-tables-on-househole
		projections
Town centres	Data: Location of town centres	Data: Number of households
	in 2004.	each output area.
	Source: MHCLG Town Centre	Source: 2011 Census + LA
	and retail planning statistics for	updates from MHCLG mid-
	England and Wales.	year household projections of
		calculation vear.
	Further information: https://	Further information: 2011
	data.gov.uk/dataset/	Census: http://www.nomiswe
	ed07b21f-0a33-49e2-9578-	co.uk/census/2011
	83ccbc6a20db/english-town-	
	centres-2004	MHCLG mid-year household
		projections: <u>https://www.gov.</u>
	U	uk/government/statistical-dat
		sets/live-tables-on-household
		projections

http://bmjopen.bmj.com/

/site/about/gu

2 3

4

5

6 7

8

9

48

57

58

GP destination data

The GP surgery destinations used from 2014 to 2016 are based on the list of practices maintained by the Organisational Data Service of the Health & Social Care Information Centre, and published at https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-

related-data. This was supplemented with information on branch surgeries from the same source. Grid references were derived from the postcode using the Office for National Statistics (ONS)

Postcode Address File. Practices with identical postcodes were taken to be duplicates or colocated, and all additional records after the first were removed. 10

11 From 2017, the list of GP locations is taken from the NHS Digital publication of Registered patients 12 at GP practices for October of the calculation year. This had the effect of reducing the number 13 14 of locations in the dataset, but removed the need for manual adjustments and produces a more 15 stable list defined as GP practices with registered patients. Grid references were derived from the 16 postcode using the Office for National Statistics (ONS) Postcode Address File. 17

18 Hospital destination data 19

20 The starting point for hospital sites is the Care Quality Commission's (CQC) list of 'active locations' 21 dataset, which is thought to be the most-up-to date and freely available source of data on individual 22 23 National Health Service (NHS) and social care 'sites' or hospitals. A criteria was developed in 24 consultation with the Department of Health to reduce the list down to capture only the key hospitals. 25 The following have been removed and individual records have been inspected to remove further 26 27 examples of these cases and for any duplicates: 28

- 29 care home records; •
- 30 non-NHS providers; 31
- 32 sites not associated with acute providers; ٠
- 33 any remaining sites that are associated with Specialist Trusts (usually single speciality Trusts or • 34 35 Sites):
- 36 records where it is evident from the name that the record is not a hospital (e.g. headquarters, 37 specialist units.) 38

39 This gave a final list of 278 hospitals in 2017 run by Acute (non-specialist) Trusts. As well as 40 covering all general hospitals this will still include some with a largely or entirely community or 41 42 rehabilitation role, where these happen to be managed by an Acute Trust. It was considered on 43 balance better to leave these in the list, rather than risk adding further subjectivity to the selection. 44 Whilst not perfect, it is considered that the resulting list is a significant improvement on that used 45 46 previously. 47

Steps taken to produce hospital data set

49 Remove records where Care Home = Y 50

51 Remove records where Provider ID begins 1-52

53 Keep records where **Benchmark Group** is Care Home or **Cluster Group** is Acute 54

Filter the trust site locations by name to remove obvious non-hospital sites. Key words 55 56

used for this process are: birth, dental, house, clinic, grange, lodge, infirmary, health, community, unit, surgery, centre

59 Manual review of remaining locations 60

Journey Time Statistics: Notes and Definitions - Page 13

Employment destination data

The employment centres are defined by the number of jobs existing in each English LSOA, taken from the Business Register Employment Survey. Large Employment Centres are defined as those with 5,000 or more jobs, Medium Employment Centres as those with 500 or more jobs, up to 4,999 and Small Employment Centres as those with 100 or more jobs, up to 499.

Data are downloaded from the Nomis website; although LSOA level BRES data has safeguarded
 access, access can be requested through the site. The chosen data download options are
 LSOA2011 geography, date as calculation year, variable as employment status where the value is
 employed, and the measure chosen is a count.

For the 2016 destination set, the BRES changed from 2001 census geography to 2011 census geography. The majority of LSOA boundaries are unchanged between these datasets, but some have been merged or split. Therefore the employment destination indicators are not strictly comparable between 2015 and 2016 Journey Time statistics. See <u>https://www.ons.gov.uk/</u> <u>methodology/geography/ukgeographies/censusgeography</u> for further information.

Education destination data

The education destination datasets are taken from the Department for Education database of educational establishments. The database was filtered to remove those establishments that were not open during the school year starting in September of the calculation year. Further filters were applied to remove special educational establishments, boarding schools and selective schools, and then to select schools at each phase of education for primary and secondary schools and further educational establishments. The following table lists the filters used.

iez oni

Phase of	Code Variable	Variable	Selec	ted codes and values
Education				
All Schools	OpenDate			30/08/17 or earlier; NUL
	CloseDate	1		30/08/18 or later; NULL
	TypeOfEstablishment_	TypeOfEstablishment	1	Community school
			2	Voluntary aided schoo
			3	Voluntary controlled
			-	school
			5	Foundation school
			6	City technology collect
			12	Foundation special
				school
			18	Further education
			28	Academy sponsor led
			29	Higher education
		N.		institutions
			31	Sixth form centres
			32	Special post 16
				institution
			34	Academy converter
			35	Free schools
			36	Free schools special
		14	39	Free schools 16 to 19
			40	University technical
		C		college
			41	Studio schools
			45	Academy 16-19
				converter
			46	Academy 16 to 19
				sponsor led
	Boarders_Code_	Boarders	0	Not applicable
			1	No boarders
			9	NULL
	AdmissionsPolicy_Code_	AdmissionsPolicy	0	Not applicable
			4	Non-selective
			9	NULL
	PhaseOfEducation Code	PhaseOfEducation	2	Primary
Primary				.

Journey Time Statistics: Notes and Definitions - Page 15

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Phase of	Code Variable	Variable	Selec	ted codes and values
Secondary	PhaseOfEducation Code		0	Not applicable
		ThaseOrEducation		
schools			4	Secondary
			5	Middle deemed secondary
			7	All through
	Statutory High age		>=16	
	Statutory Low age		< 16	
FE	PhaseOfEducation_Code_	PhaseOfEducation	4	Secondary
			5	Middle deemed secondary
			6	16 plus
			7	All through
	Statutory High age	·	>16	
	OfficialSixthForm_Code_	OfficialSixthForm	0	Not applicable
			1	Has a sixth form
		9	NULL	
		OR		
FE	EstablishmentTypeGroup	EstablishmentTypeGroup	1	Colleges
	code_			

Food Stores destination data

The food stores destination dataset is purchased from <u>The Local Data Company</u> and includes all branches of multiple food store chains. Although some data are available for independent food stores, this only exists within town centres and so has not been included.

Connectivity

Destinations	Data source for the locations	Data source for users of the
Destinations Airports	Data source for the locations of the service Data: Location of GB airports excluding highlands and islands of Scotland Source: National Public Transport Access Nodes Further information: https:// data.gov.uk/dataset/ ff93ffc1-6656-47d8-9155- 85ea0b8f2251/national-public- transport-access-nodes-naptan	Data source for users of the service Data: Number of households in each output area. Source: 2011 Census + LA updates from MHCLG mid- year household projections of calculation year. Further information: 2011 Census: <u>http://www.nomisweb.</u> <u>co.uk/census/2011</u> MHCLG mid-year household projections: <u>https://www.gov.</u>
		uk/government/statistical-data- sets/live-tables-on-household- projections

Destinations	Data source for the locations of the service	Data source for users of the service
Railway stations	Data: Location of larger (category A, B and C1) rail stations in GB Source: Network rail classification	Data: Number of households in each output area. Source: 2011 Census + LA updates from MHCLG mid- year household projections of
	Further information: <u>http://webarchive.</u> <u>nationalarchives.gov.</u> <u>uk/20101007153226/</u> http://www.dft.gov.uk/pgr/	Calculation year. Further information: 2011 Census: <u>http://www.nomisweb.</u> <u>co.uk/census/2011</u> MHCLC mid_year bousehold
	rail/passenger/stations/ betterrailstations/ http://archive.nr.co.uk/	projections: <u>https://www.gov.</u> uk/government/statistical-data- sets/live-tables-on-household-
	browse%20documents/ rus%20documents/route%20 utilisation%20strategies/	projections
	network/working%20 group%202%20-%20stations/ networkrusstations.pdf	

Transport network data

Travellers moved between their original and their destination via one or more of the following transport networks, depending on the mode of transport being modelled. For all modes, travellers will probably also need to walk between their origin / destination and the transport network. For some short journeys, it may be quicker for travellers to walk directly to their destination, rather than using public transport at all – this is why public transport / walking results are modelled as a combined mode.

Public transport

National public transport timetable data are publically available. Data for bus, local coach and other local transport services (e.g. light rail, metro, and ferry) are captured in the Traveline National Data Set (TNDS), rail timetable data are published by the Association of Train Operating Companies (ATOC), and national coach services in the National Coach Data Set (NCDS).

Walk

The walking network is represented by the road and urban path elements of the Integrated Transport Network produced by the Ordnance Survey.

Cycle

The cycling network is represented by the road network including cycle paths and bridleways from the Integrated Transport Network. Cycle journeys are also allowed to use footpaths at walking pace.
Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

Car

The car network is represented by the road component of the Integrated Transport Network.

Data on actual vehicle speeds on each road network link (generally the stretch of road between 2 nodes, or junctions) is obtained from Trafficmaster Satnav devices and are used to estimate car speeds. These data are used to calculate annual average traffic speeds on each link of the road network (by direction if the link is bi-directional). These are used as the link speeds for cars in the modelling. Where the Trafficmaster sample for an individual link is too small, national averages of the same data for the particular road type are used instead. This is an innovation from 2014. Previously the sample was too small and the model reverted to default assumptions for car speed based on road type which were much higher than the Trafficmaster averages, resulting in some inconsistency in the model.

Outputs

The journey time results are used to create the following indicators for publication:

Indicator		Description
Minimum journey time		The shortest of the ten journey time results.
Origin indicators	6	Four measures, the number of destinations (up
		to the maximum of 10) that can be reached
		from a given origin within 15, 30, 45 and 60
		minutes.
Destination indicators		Four measures, the percentages of service
		users within the given geographical area who
		can access at least one service location within
		15, 30, 45 and 60 minutes.

Each of these indicators is calculated for each mode and each destination type, and at a number of geographical scales as follows:

England

Region

► Local Authorities, including London Boroughs, Metropolitan districts, Unitary authorities, Counties and non-Metropolitan districts, also Inner and Outer London and former Metropolitan counties

- ► 2011 Lower layer Super Output Area
- ► 2011 Defra Rural/Urban Classification

The indicators for each geography are calculated as population weighted averages. In other words, the average minimum journey time for an area, B, is:

mjt(B)= ∑(i=1)^n(mjt(OAi)×pop(OA_i))/pop(B)

where mjt(B) is the minimum journey time in area B, mjt(OAi) is the minimum journey time of the ith of n output areas making up area B, and pop(B) and pop(OAi) are the user populations resident in area B and output area i respectively.

The service user populations used in the above weighting, and in the destination indicators, depend on the destination type, as follows:

Destination type	Service user population basis
Employment centres	Resident population of working age (16-74
	years)
Primary schools	Population aged 5-10
Secondary schools	Population aged 11-15
Further education colleges	Population aged 16-19
GPs, hospitals, food stores, town centres	Number of households
Average key services	Resident population of working age (16-74
	years)

Strengths and Weaknesses

In using the data, the following points should be kept in mind:

All journey times are compiled on a consistent basis across the country.

► The statistics are based on the calculation of theoretical journey times, they are not based on real journeys. They are however based on actual public transport times, and average traffic speeds on the road network.

Although the statistics are calculated to a high level of geographical detail, some assumptions and simplifications are necessary in the modelling (for example assigning the start point of journeys to a single point in each Output Area, road speeds, interchange times for public transport).

► For 2016 we have used the 2015 BRES data to designate Lower Super Output Areas as employment centres. The 2015 BRES is the first year to use LSOAs based on the 2011 census, and although the majority of these are an exact match to the 2001 LSOAs, there are some that were merged, split or had other boundary changes. For these areas journey times from earlier years are not comparable to the 2016 journey times. This effect is more pronounced for large employment centres, as there are fewer destinations to route to.

► For particular areas, local authorities and other experts may have more detailed information allowing them to produce more accurate or detailed models of the local situation.

Demand responsive services (e.g. bus services which have to be booked) are only included to the extent that they can be plausibly modelled, in the Traveline National Data Set.

► Since new journey calculation software was adopted for 2014, along with a significant number of other changes to the methodology, from 2014 results are not directly comparable with those for earlier years.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

3
4
5
6
7
/ 0
8
9
10
11
12
13
14
15
16
17
17
18
19
20
21
22
23
24
25
26
20
27
28
29
30
31
32
33
34
35
36
20
3/
38
39
40
41
42
43
44
15
75 76
40
4/
48
49
50
51
52
53
54
55
55
56
57
58
59
60

Supplementary material S1 – OPCS-4	code criteria	used for	Hospital	Episode
Statistics data extraction				

Code	Code description
OPCS-4 cc	odes for knee revision procedures
O180	Conversion from previous hybrid prosthetic replacement of knee joint using cement
O182	Conversion to hybrid prosthetic replacement of knee joint using cement
O183	Revision of hybrid prosthetic replacement of knee joint using cement
O184	Attention to hybrid prosthetic replacement of knee joint using cement
W400	Conversion from previous cemented total prosthetic replacement of knee joint
W402	Conversion to total prosthetic replacement of knee joint using cement
W403	Revision of total prosthetic replacement of knee joint using cement
W404	Revision of one component of total prosthetic replacement of knee joint using cement
W410	Conversion from previous uncemented total prosthetic replacement of knee joint
W412	Conversion to total prosthetic replacement of knee joint not using cement
W413	Revision of total prosthetic replacement of knee joint not using cement
W414	Revision of one component of total prosthetic replacement of knee joint not using cement
W420	Conversion from previous total prosthetic replacement of knee joint NEC
W422	Conversion to total prosthetic replacement of knee joint NEC

W423	Revision of total prosthetic replacement of knee joint NEC
W424*	Attention to total prosthetic replacement of knee joint NEC
W425	Revision of one component of total prosthetic replacement of knee joint NEC
W522†	Conversion to prosthetic replacement of articulation of bone using cement NEC
W523†	Revision of prosthetic replacement of articulation of bone using cement NEC
W532†	Conversion to prosthetic replacement of articulation of bone not using cement NEC
W533†	Revision of prosthetic replacement of articulation of bone not using cement NEC
W542†	Conversion to prosthetic replacement of articulation of bone NEC
W543†	Revision of prosthetic replacement of articulation of bone NEC
W544*†	Attention to prosthetic replacement of articulation of bone NEC
W553†	Conversion to prosthetic interposition arthroplasty of joint
W564†	Conversion to interposition arthroplasty of joint NEC
W574†	Conversion to excision arthroplasty of joint
W582†	Revision of resurfacing arthroplasty of joint
W603†	Conversion to arthrodesis and extra-articular bone graft NEC
W613†	Conversion to arthrodesis and articular bone graft NEC
W641†	Conversion to arthrodesis and internal fixation NEC
W642†	Conversion to arthrodesis and external fixation NEC
OPCS-4 co	odes for laterality

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

3
4
5
6
7
, 0
ð
9
10
11
12
13
14
15
16
17
10
18
19
20
21
22
23
24
25
26
20
27
28
29
30
31
32
33
34
35
36
20
3/
38
39
40
41
42
43
44
45
75 76
40
4/
48
49
50
51
52
53
54
55
22
50
57
58
59
60

1 2

Z942	Left-sided
Z943	Right-sided
ICD-10 co	odes for Infection
T845	Infection and inflammatory reaction due to internal joint prosthesis
T846	Infection and inflammatory reaction due to internal fixation device [any site]
T847	Infection and inflammatory reaction due to other internal orthopaedic prosthetic devices, implants and grafts
T814	Infection following a procedure, not elsewhere classified
ICD-10 co	odes for fracture
M966	Fracture of bone following insertion of orthopaedic implant, joint
	prosthesis or bone plate
ICD-10 co	odes for mechanical complications
T840	Mechanical complication of internal joint prosthesis
T841	Mechanical complication of internal fixation device of bones of limb
T842	Mechanical complication of internal fixation device of other bones
T843	Mechanical complication of other bone devices, implants and grafts
T844	Mechanical complication of other internal orthopaedic devices,
	imnplants and grafts
ICD-10 co	odes for osteoarthritis/arthrosis
M15-	Polyarthrosis
M17-	Gonarthrosis
M19-	Other arthrosis
OPCS-4 =	Office of Populations Censuses and Surveys Classification of
Interventio	ns and Procedures version 4. ICD-10 = International Statistical

Classification of Diseases and Related Health Problems, tenth revision. * Where

OPCS-4 codes Y032 (renewal of prosthesis in organ NOC) or Y037 (removal of prosthesis from organ NOC) were also used. † Where OPCS-4 codes O132 (knee NEC) or Z765 (lower end of femur NEC) or Z774 (upper end of tibia NEC) or Z787 (patella) or Z844 (patellofemoral joint) or Z845 (tibiofemoral joint) or Z846 (knee joint) or Z851 (upper tibiofibular joint) were used to identify knee as the body site.

for perteries only





77x73mm (300 x 300 DPI)

BMJ Open

Supplementary material S4 – R Code	
#Travel Times and Perioperative Outcomes in Revision Knee Replacement	
setwd("/Users/alexandermatthews//OneDrive - University of Exeter/Alex Mat MD/Revision Knee Networks MD/Travel Times Analysis_/")	thews
####Preparation of Data#### #load HES data	
RTKA2023 <- read.csv("~/Desktop/RTKA 06-09-23 CSV.csv")	
RTKA2023 <- read.csv("/Users/alexandermatthews//OneDrive - University of E Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/RTKA 06-0 CSV.csv")	xeter/Alex 19-23
#table only shows first 50 columns but we know there are 51 columns. Write t code to change preferences	his generic
rstudioapi::writeRStudioPreference("data_viewer_max_columns", 1000L)	
#Some entried are blank but are read as real values and not missing data #The table between age and sex shows three variables here #The dataset contains non standard missing values that are not recognised as #Replace empty strings with NA	NA
RTKA2023[RTKA2023 == ""] <- NA	
#Find number of incomplete cases in the data	
missing_data <- colSums(is.na(RTKA2023)) print(missing_data)	
#There are 14 entries with missing data only in the age group	
#check how many incomplete entries in age of patient column	
<pre>sum(!complete.cases(RTKA2023\$age_of_patient))</pre>	
#In case of missing values there are only 14 for age of patient #Can use imputation based on mean age #What is the mean age of the patients	

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

3
1
4
5
6
7
8
a
10
10
11
12
13
14
15
16
17
17
18
19
20
21
22
22
23
∠4 25
25
26
27
28
29
30
21
31
32
33
34
35
36
37
20
38
39
40
41
42
43
44
 1 E
45
46
47
48
49
50
51
51
52
53
54
55
56
57
52
50
59
60

1 2

mean(RTKA2023\$age_of_patient, na.rm = TRUE)

#mean age excluding missing values is 70
summary(RTKA2023\$age_of_patient, na.rm = TRUE)

#Check age is normally distributed

hist(RTKA2023\$age_of_patient)

#Input mean for missing values for age

RTKA2023\$age_of_patient[is.na(RTKA2023\$age_of_patient)] <- 69.82

#Now check number of missing values

sum(!complete.cases(RTKA2023\$age_of_patient))
#Now states 0 missing values

#There are other missing values for IMD decile ##In fact there are 439 IMD score missing values

sum(!complete.cases(RTKA2023\$IMD_score))

hist(RTKA2023\$IMD_score) #IMD score is non normally distributed

summary(RTKA2023\$IMD_score, na.rm = TURE)

#Median IMD score is 15.543

#Use imputation to impute median for missing value

RTKA2023\$IMD_score[is.na(RTKA2023\$IMD_score)] <- 15.543

#Check imputation complete

sum(!complete.cases(RTKA2023\$IMD_score))

#Now showing 0 missing values

#Next attach IMD decile number 6 to the missing values. As a score of 15 equates to the 6th decile

RTKA2023\$IMD_decile[is.na(RTKA2023\$IMD_decile)] <- 6

3	
4	
т 5	
5	
0	
/	
8	
9	
10	
11	
12	
13	
14	
15	
16	
10	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
20	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
20	
20	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
10	
47 50	
50	
51	
52	
53	
54	
55	
56	
57	
58	
50	
22	
οU	

#Check duplicate entry spells

duplicates <- RTKA2023[duplicated(RTKA2023),]

#No duplicates in data

#Frequencies of revisions by volume

as.numeric(RTKA2023\$TV12mo)

#frequencies of revisions by trust volume table(RTKA2023\$TVcat)

#Proportions by trust volume

prop.table(table(RTKA2023\$TVcat))

#Some entried are blank but are read as real values and not missing data #The table between age and sex shows three variables here #The dataset contains non standard missing values that are not recognised as NA #Replace empty strings with NA Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

RTKA2023[RTKA2023 == ""] <- NA

#Check this has registered

missing_data <- colSums(is.na(RTKA2023))
print(missing data)</pre>

#Column with LSOA_2011_Code has 171 missing.

#LSOA is part of primary exposure variable, small number of missing cases. Decision to remove rows rather than estimate from imputation because factor variable and dependent on provider code. Multiple imputation was used later to estimate missing travel data for these multiple rows where LSOA and site code was available

#Remove missing data in dataframe combined_data for column LSOA_2011_Code with missing fields = 171

RTKA2023<- RTKA2023[!is.na(RTKA2023\$LSOA_2011_Code),]

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

#16,565 patients before link with TRACC travel data

#Load Travel times data

TRAVELTIMES <- read.csv("~/Desktop/Drive time and Miles reference file.csv")

LSOAREF <- read.csv("~/Desktop/LSOA Matrix.csv")

LSOAREF <- read.csv("/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/LSOA Matrix.csv")

#Join data but The data is too big so we need to do this using SQL

install.packages("RSQLite") library(RSQLite)

query <- " Select * FROM times JOIN Isoa ON times.LSOAName = Isoa.LSOA11NM"

result <- dbGetQuery(con, query)

#10million 457 thousand and 999 possible combinations

#Write Dataframes

write.csv(result, "~/Desktop/JOINLSOATRAVEL.csv")

result<- read.csv("/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/JOINLSOATRAVEL.csv")

#####Now join this data to your revisions spreadsheet using key identifiers LSOA and Organisation site code

C	query <- "
ç	Select *
F	FROM revisions3
J	IOIN travel3 ON revisions3.LSOA_2011_Code = travel3.LSOA11CD AND revisions3.Sitecode =
t	ravel3.ProviderSiteCode"
r	result_join <- dbGetQuery(con, query)
ŧ	#Number of patients following join 12,774
r	result1 <- result_join
Ŧ	Check your data for missing values
r	missing_data <- colSums(is.na(result1))
F	print(missing_data)
\$	#Check data for duplicates
(duplicates <- RTKA2023[duplicated(RTKA2023\$Epikey),]
‡	# Check for duplicates in the 'epikey' column
(duplicates <- result1[duplicated(result1\$Epikey),]
‡	#There are 2,047 duplicates
‡	#Remove duplicates in result 1
‡	# Remove duplicates: Keep only the first occurrence of each 'Epikey'
r	result1 <- result1[!duplicated(result1\$Epikey),]
‡	#final dataframe is 10,727
١	write.csv(result1, "/Users/alexandermatthews//OneDrive - University of Exeter/Alex
ſ	Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/FinalJOIN.csv")
‡	####Prepare Outcomes, Exposure variable and co-variates ####
,	*Cot up outcomos
1	rset up outcomes
1	

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

```
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

> result1\$Read30 <- ifelse(is.na(result1\$Read30), 'N', result1\$Read30) result1\$Read90 <- ifelse(is.na(result1\$Read90), 'N', result1\$Read90)

```
result1$Read30days <- ifelse(result1$Read30 == "Y", 1, 0)
#readmission for 90 days
result1$Read90days <- ifelse(result1$Read90 == "Y", 1, 0)</pre>
```

#Set up your co-variates

```
result1$HFRS_Band = as.factor(result1$HFRS_Band)
result1$HFRS_Band = relevel(result1$HFRS_Band, ref = 'None')
```

result1\$POD = as.factor(result1\$POD)
result1\$POD = relevel(result1\$POD, ref = 'EL')

table(result1\$POD)

#I've joined two dataframes based on a shared field. But some rows have not jointed

#Journey times statistics - 10,457,999 rows

#12,774 following join with revisions and travel data called "result1" but had duplicates 2,047 so remove these (duplicates due to slightly different latitude and longitude for same Site codes in journey times statistics)

#Final results 1 following removal of duplicates is 10,727

#Original dataframe is 16,736 called RTKA2023 following removal of early revisions, excluding missing LSOA was 16565

#Missing data for travel seen in 5,838 patients or 35% of patients

#Use multiple imputation to impute missing distance values for cases without join

#How many unmatched rows?

unmatched_rows <- RTKA2023[!(RTKA2023\$Epikey %in% result1\$Epikey),]

#There are 5,838 unmatched rows

#I want to create a dataframe showing both matched and unmatched fields based on this.

Identify columns that are in result1 but not in RTKA2023
missing_cols <- setdiff(names(result1), names(RTKA2023))</pre>

Add missing columns to RTKA2023 with NA values for (col in missing_cols) { RTKA2023[[col]] <- NA }
Ensure column order is the same as result1 RTKA2023 <- RTKA2023[, names(result1)]
Identify unmatched rows unmatched_rows <- RTKA2023[!(RTKA2023\$Epikey %in% result1\$Epikey),]
Combine matched rows (result1) with unmatched rows combined_data <- rbind(result1, unmatched_rows)
duplicates <- combined_data[duplicated(combined_data\$Epikey),]
#0 duplicates
write.csv(combined_data, "/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/FinalJOINCombined.csv")
combined_data <- read.csv("/Users/alexandermatthews//OneDrive - University of
Analysis_/FinalJOINCombined.csv")
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30)
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30) combined_data\$Read30days <- ifelse(combined_data\$Read30 == "Y", 1, 0)
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30) combined_data\$Read30days <- ifelse(combined_data\$Read30 == "Y", 1, 0) #Now have dataframe displaying both matched and unmatched rows
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30 combined_data\$Read30days <- ifelse(combined_data\$Read30 == "Y", 1, 0) #Now have dataframe displaying both matched and unmatched rows missing_data <- colSums(is.na(combined_data)) print(missing_data)
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30) combined_data\$Read30days <- ifelse(combined_data\$Read30 == "Y", 1, 0) #Now have dataframe displaying both matched and unmatched rows missing_data <- colSums(is.na(combined_data)) print(missing_data) #How many patients in high volume centres >49
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30) combined_data\$Read30days <- ifelse(combined_data\$Read30 == "Y", 1, 0) #Now have dataframe displaying both matched and unmatched rows missing_data <- colSums(is.na(combined_data)) print(missing_data) #How many patients in high volume centres >49 combined_data\$MRC <- ifelse(combined_data\$TV12mo > 49, 1, 0)

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

2	
3	
4	
5	
6	
0	
7	
8	
9	
10	
10	
11	
12	
13	
14	
14	
15	
16	
17	
10	
18	
19	
20	
21	
22	
22	
23	
24	
25	
25	
26	
27	
28	
20	
29	
30	
31	
32	
22	
22	
34	
35	
36	
27	
57	
38	
39	
40	
/1	
41	
42	
43	
44	
15	
45	
46	
47	
48	
⊿0	
77	
50	
51	
52	
52	
22	
54	
55	
56	
55	
5/	
58	
59	
<u> </u>	

1

nopatients <- subset(combined_data, MRC == 1)</pre> #6880 patients missing_data <- colSums(is.na(nopatients))</pre> print(missing data) # Count unique levels of ProvCode n_levels <- length(unique(nopatients\$ProvCode))</pre> cat("Number of unique providers (ProvCode):", n_levels, "\n") #38 providers #How many sites # Count unique levels of ProvCode n levels <- length(unique(nopatients\$Sitecode))</pre> cat("Number of unique sites (Sitecode):", n_levels, "\n") #187 sites #rates of readmission 30 days table(nopatients\$Read30days) #568/6880 8.3% #rates of mortality at 90 days table(nopatients\$Mort90days) #217/6880 3.2% #Rates of length of stay above median. Remember median calculated across entire cohort summary(combined data\$Spell Los) #Median of 5 nopatients\$Long_Los <- ifelse(nopatients\$Spell_Los > 5, 1, 0) table(nopatients\$Long_Los) #3421/6880 49.7% #3157 travel data not available #16,565 observations in entire dataframe not limited to teriatry referral centres #CV12mo missing 71 cases. Imputation using median due to positive skew

2	
3	
5	hist(combined_data\$CV12mo)
б	
7	#mean age excluding missing values is 70
8	summary(combined_data\$CV12mo, na.rm = TRUE)
9	
10	
12	#Input median of 6 for missing data
13	
14	combined data\$CV12mo[is.na(combined data\$CV12mo)] <- 6
15	
10	#Now need to use multiple imputation method to estimate travel data for columns
18	"DistanceMiles" "OffPeakDriveDistanceMiles" "PeakDriveTimes' based on associated
19	predictors:
20	
21	#Pofor to this resource "https://bookdown.org/mwhovmons/bookmi/multiplo-
22	imputation html#sotting the imputation methods"
23	imputation.num#setting-the-imputation-methods
24	HAndahis second for each at
26	#And this resource for context
27	https://dept.stat.lsa.umich.edu/~jerrick/courses/stat/01/notes/mi.html
28	
29	# https://www.ebpi.uzh.ch/dam/jcr:dc0cef17-29c7-4e61-8d33-
30 31	e690561ab7ae/mi_intro20191001.pdf (Advice on multi level modelling and imputation)
32	
33	# Install packages if they are not already installed
34	install.packages(c("mice", "ggplot2", "naniar"))
35	
36	# Load the packages
38	library(mice)
39	library(ggplot2)
40	library(naniar)
41	
42	#assuming missing data is due to random chance, LSOA and SiteCode are related to the
43	exposure but also include all other variables linked to your analysis
44 45	#Subset dataframe called combined date with only with relevant columns; age of patient.
46	sex HERS Band IMD Score, IMD Decile, infection, TVcat, CVcat, SiteCode, ProvCode, EinY,
47	DistanceMiles OffPeakDriveDistanceMiles PeakDriveTime Mort90days Read30 Snell Los
48	#decision not to include site code and LSOA as likely not present in missing data
49	"ISOA 2011 Codo" "Sitocodo"
50	LSOA_2011_Code, Silecode
57	
53	
54	# Constitution relevant as how as the start TMO as a set of the last to the
55	# Specify the relevant columns live included 1V12mo as may be related to outcome,
56	ProvCode for clustering,
5/	relevant_columns <- c(
50 59	"age_ot_patient", "sex", "HFRS_Band", "IMD_score",
60	"infection", "TV12mo", "CV12mo", "ProvCode", "FinY",

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

2
3
4
-
5
6
7
8
0
9
10
11
12
12
15
14
15
16
17
17
18
19
20
21
21
22
23
24
25
25
26
27
28
20
29
30
31
32
22
22
34
35
36
27
57
38
39
40
⊿1
41
42
43
44
45
45
46
47
48
10
49
50
51
52
52
55
54
55
56
57
57
58
59
60

1

"DistanceMiles", "OffPeakDriveDistanceMiles", "PeakDriveTime", "Mort90days", "Read30days", "Spell_Los"

)

Subset the dataframe with only the relevant columns subset_combined_data <- combined_data[, relevant_columns]</pre>

#Currently sex, HFRS_Band, TVCat, Sitecode, ProvCode, FinY are not incorporated in model as character variables

#convert these to factors

Convert variables to factors
subset_combined_data\$sex <- as.factor(subset_combined_data\$sex)
subset_combined_data\$ProvCode <- as.factor(subset_combined_data\$ProvCode)
subset_combined_data\$FinY <- as.factor(subset_combined_data\$FinY)
subset_combined_data\$HFRS_Band <- as.factor(subset_combined_data\$HFRS_Band)</pre>

subset_combined_data\$Sitecode <- as.factor(subset_combined_data\$Sitecode)
subset_combined_data\$LSOA_2011_Code <as.factor(subset_combined_data\$LSOA_2011_Code)</pre>

Check the structure of the dataframe to confirm str(subset_combined_data[, c("sex", "Sitecode", "ProvCode", "FinY", "HFRS_Band", "LSOA_2011_Code")])

#visualise missing data

vis_miss(subset_combined_data)

#35% missing travel data

Set the seed for reproducibility set.seed(123)

Perform Multiple Imputation

imp <- mice(subset_combined_data, m=5, method='pmm')</pre>

#Check for imputation values

imp\$imp\$OffPeakDriveDistanceMiles

#visualise imputed values
imp\$imp
#Means of the imputed values
imp\$chainMean
#What are the predictors
imp\$predictorMatrix
#Plot imputation values against observed values.
my_plot <- stripplot(imp, col=c("grey", "blue"), pch = c(1, 20))
my_plot
#Guidelines for imputation model suggest all variables in the analysis should be included, inclusive of dependent or outcome variables
#Ensure TVCat is not a predictor variable
pred <-imp\$predictorMatrix pred["TVcat"] <- 0 pred
#Plot the convergence (how equal is the variance to the mean)
plot(imp)
#Stack the imputed values into a single dataset and include original data
imp2 <- complete(imp, "long", inc = TRUE)
#Save imp2
write.csv(imp2, "/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/imp2.csv")
#Read it back in here:
imp2 <- read.csv("/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/imp2.csv")

BMJ Open

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

#Save as Supplemenatry figure

#Filter data by tertiary hospitals only

#But current guidelines suggest >49 is a high volume centre called a major revision centre and probably represents a unit with tertiary specialisation imp2\$MRC <- ifelse(imp2\$TV12mo > 49, 1, 0) tertiary revisions <- subset(imp2, MRC == 1)</pre> tertiary revisions\$Long Los <- ifelse(tertiary revisions\$Spell Los > 5, 1, 0) #declare the imputed data to be mids again, the format MICE is expecting for regression analyses tertiary revisions <- as.mids(tertiary revisions) #Now run your regression model using a multivariable model #A priori co-variates chosen based on evidence of predictors for readmission ####Primary Outcome 30 day readmission #### #Exposure 1 - Distance Miles library("lme4") # Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for clustering m3.mi <- with(tertiary revisions, glm(Read30days ~ DistanceMiles + IMD score + HFRS Band + sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode, family = "binomial")) print(m3.mi) # Pool results across imputed datasets pooled results <- pool(m3.mi) # Summarize pooled results with confidence intervals

Summarize pooled results with confidence intervals
summary_pooled <- summary(pooled_results, conf.int = TRUE)</pre>

Add Odds Ratios to the summary summary_pooled\$OR <- exp(summary_pooled\$estimate) summary_pooled\$Lower_CI <- exp(summary_pooled\$`2.5 %`) summary_pooled\$Upper_CI <- exp(summary_pooled\$`97.5 %`)
Display the final table with Odds Ratios and Confidence Intervals print(summary_pooled)
#check for evidence of multicollinearity?
library(car)
Use the long data including all imputations for VIF
<pre>tertiary_revisions <- complete(tertiary_revisions, "long", inc = TRUE)</pre>
<pre># Fit a logistic regression model on the complete dataset vif_model <- glm(Read30days ~ DistanceMiles + IMD_score + HFRS_Band + sex + age_of_patient + infection + TV12mo + CV12mo + FinY + ProvCode, data = tertiary_revisions, family = "binomial")</pre>
<pre># Calculate VIF vif_values <- vif(vif_model) print(vif_values)</pre>
#No evidence of multi-collinearity
#Is there a non linear relationship?
#Box Tidwell
#Recode back into correct format
tertiary_revisions <- as.mids(tertiary_revisions)
<pre># Custom function to add log-transformed variable and interaction term add_interaction <- function(data) { data\$Log_DistanceMiles <- log(data\$DistanceMiles) # Add log-transformed variable data\$Interaction <- data\$DistanceMiles * data\$Log_DistanceMiles # Add interaction term return(data) }</pre>

# Extract the long-format data including the original data	
tertiary_revisions_modified <- complete(tertiary_revisions, action = "long", include = TRUE)
# Apply the transformation to each imputed dataset	
tertiary_revisions_modified <- do.call("rbind",	
lapply(split(tertiary_revisions_modified,	
tertiary_revisions_modified\$.imp),	
# Convert back to mids object	
tertiary_revisions_modified <- as.mids(tertiary_revisions_modified)	
# Fit the logistic regression model with the interaction term	
model <- with(tertiary_revisions_modified, glm(Read30days ~ DistanceMiles + Interaction,	
data = tert	
Tariniy = binornal(intk = Togit)))	
# Pool the results	
pooled_results <- pool(model)	
# Summarize pooled results	
summary_pooled <- summary(pooled_results, conf.int = TRUE)	
# Extract the p-value for the interaction term	
box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]	
print the p-value	
p(p)	
# p value = 0.03 evidence of non linearity	
#Are spline terms significant for DistanceMiles if using 3 knots, 4 knots and 5 knots	
#Use data of all imputations in long format	
tertiary_revisions <- complete(tertiary_revisions, "long", inc = TRUE)	
# Load the required library	
library(splines)	
#AIC of non spline model	
model <- glm(Read30days ~ DistanceMiles, data = tertiary_revisions, family = binomial)	
Summary(model)	

3	
4	HAIC 21862
5	#AIC 21802
6	
7	# Define a function to fit and evaluate spline models with knots based on centiles
8	evaluate centile splines <- function(centiles, data) {
9	# Calculate knots based on the specified contiles
10	
11	knots <- quantile(data\$DistanceMiles, probs = centiles, na.rm = TRUE)
12	
13	# Fit a logistic regression model with natural splines using the calculated knots
14	model_spline <- glm(Read30days ~ ns(DistanceMiles_knots = knots)
15	family = hinamial/link = "logit")
16	Tamity = binomial(link = Togit),
17	data = data)
18	
19	# Summarize the model
20	summary model <- summary(model_spline)
21	summary_model v summary(model_spinley
22	
23	# Extract p-values for the spline terms
24	p_values <- summary_model\$coefficients[-1, "Pr(> z)"] # Exclude the intercept
25	
26	# Print the results
27	cat("\nRosults for contilos" contilos ":\n")
28	cat (intestits for centiles, centiles,)
29	print(p_values)
30	
31 22	# Return the model and calculated knots for further inspection if needed
2Z 22	return(list(model = model_spline, p_values = p_values, knots = knots))
37	}
35	
36	# Evenue of the configurations for 2.4 and 5 linets
37	# Example centile configurations for 3, 4, and 5 knots
38	centiles_3_knots <- c(0.05, 0.50, 0.95) # 5th, 50th, and 95th percentiles
39	centiles_4_knots <- c(0.05, 0.35, 0.65, 0.95) # Custom centiles for 4 knots
40	centiles 5 knots <- c(0.05, 0.25, 0.50, 0.75, 0.95) # 5 knots centiles
41	
42	# Evaluate models with contile based knots using your detect
43	
44	results_3_knots <- evaluate_centile_splines(centiles = centiles_3_knots, data =
45	tertiary_revisions)
46	results_4_knots <- evaluate_centile_splines(centiles = centiles_4_knots, data =
47	tertiary revisions)
48	results 5 knots <- evaluate centile splines(centiles - centiles 5 knots data -
49	testion revisions)
50	
51	
52	# Compare models with centile-based knots
53	cat("\nComparing models with different centile-based knots:\n")
54	anova(results 3 knots\$model, results 4 knots\$model, results 5 knots\$model, test =
55	"Chica")
50	
5/ E0	
50	# Print the calculated knot locations for each model
60	cat("\nKnot locations for 3 knots:\n")
~ ~	

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

print(results 3 knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results 4 knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results 5 knots\$knots) #AIC better fit 21806 #Model with 3 knots, significant terms but greater knots do not improve the model fit. Non linear relationship is evident and should be modelled with splines #Prepare predictors for model prediction #you need to ensure that the predicted probabilities align with the corresponding observations #Explore the data for missing values sum(!complete.cases(tertiary revisions\$DistanceMiles)) #Unimputed dataset is missing, so exclude these tertiary_revisions <- tertiary_revisions[!is.na(tertiary_revisions\$DistanceMiles),] sum(!complete.cases(tertiary_revisions\$sex)) sum(!complete.cases(tertiary revisions\$Read30days)) sum(!complete.cases(tertiary_revisions\$HFRS_Band)) sum(!complete.cases(tertiary revisions\$IMD score)) sum(!complete.cases(tertiary revisions\$infection)) #Currently infection as numeric - ensure is factor tertiary revisions\$infection <- as.factor(tertiary revisions\$infection) tertiary revisions\$HFRS Band <- as.factor(tertiary revisions\$HFRS Band) tertiary revisions\$sex <- as.factor(tertiary revisions\$sex)</pre> tertiary_revisions\$FinY <- as.factor(tertiary_revisions\$FinY)</pre> tertiary revisions\$ProvCode <- as.factor(tertiary revisions\$ProvCode) tertiary_revisions\$DistanceMiles <- as.numeric(tertiary_revisions\$DistanceMiles) tertiary revisions\$age of patient <- as.numeric(tertiary revisions\$age of patient) tertiary revisions\$IMD score <- as.numeric(tertiary revisions\$IMD score)

tertiary revisions\$TV12mo <- as.numeric(tertiary revisions\$TV12mo)

BMJ Open

2	
3	tertiary_revisions\$CV12mo <- as numeric(tertiary_revisions\$CV12mo)
4	
5	
6	#Run spline model with adjusted data excluding missing data
7	library(splines)
8	# For example, let's say you want 3 knots at specific percentiles
9	knots <- quantile(tertiary revisions\$DistanceMiles probs = $c(0.05, 0.50, 0.95)$ na rm =
10	$T_{\text{D}}(\mathbf{r}) = c(0.05, 0.50, 0.50), the set of t$
11	
12	print(knots)
13	#Knots at 53, 69 and 84
14	spline terms <- ns(tertiary revisions\$DistanceMiles, knots = knots)
15	
16	
17	
18	
19	model_with_custom_splines <- glm(Read30days ~ ns(DistanceMiles, knots = knots) +
20	HFRS Band + IMD score +
21	
22	family = "binomial" data = tortiary rovisions)
23	ranniy – binonnar, uata – tertiary_revisions)
24	
25	
20	summary(model_with_custom_splines)
27	
20	#Generate a sequence of mean unit values for predicting
30	#Generate a sequence of mean unit values for predicting
31	
37	DistanceMiles_range <- seq(min(tertiary_revisions\$DistanceMiles),
32	max(tertiary revisions\$DistanceMiles), length.out = 100)
34	
35	new data <- expand grid(
36	Distance Miles - Distance Miles range
37	Distanceivilles = Distanceivilles_range,
38	sex = levels(tertiary_revisions\$sex), # Ensure it takes all factor levels
39	age_of_patient = mean(tertiary_revisions\$age_of_patient, na.rm = TRUE),
40	HFRS Band = levels(tertiary revisions\$HFRS Band), # Ensuring correct factor levels
41	IMD score = mean(tertiary revisionsSIMD score narm = TRUE)
42	FinV = lovels(tertiany_revisions\$1115_seere, name = neer,
43	rint – levels(lertiary_revisions; rint), # Ensuring correct factor levels
44	CV12mo = mean(tertiary_revisions\$CV12mo, na.rm = TRUE),
45	TV12mo = mean(tertiary_revisions\$TV12mo, na.rm = TRUE),
46	ProvCode = levels(tertiary_revisions\$ProvCode), # Ensuring correct factor levels
47	infection = levels(tertiary revisions\$infection) # Ensuring correct factor levels
48)
49	
50	
51	# Create a new dataset with a range of distances and miles and all other predictor variables
52	new_data <- expand.grid(DistanceMiles = DistanceMiles_range,
53	sex = unique(tertiary revisions\$sex),
54	age of patient = mean(tertiary revisions sage of natient)
55	HERC Rand - uniquo/tortiary_revisions¢HERC Rand)
50	IIIRS_Dallu – ullique(tertiary_revisiolispiners_Dallu),
5/	IND_score = mean(tertiary_revisions\$IMD_score),
20 20	FinY = unique(tertiary_revisions\$FinY),
59 60	CV12mo = mean(tertiary revisions\$CV12mo),
00	· · · · · · · · · · · · · · · · · · ·

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

TV12mo = mean(tertiary_revisions\$TV12mo), infection = unique(tertiary_revisions\$infection))
Align the levels of ProvCode in new_data to match the training data new_data\$ProvCode <- factor(new_data\$ProvCode, levels = levels(tertiary_revisions\$ProvCode))
<pre># Align the levels of all relevant categorical variables new_data\$HFRS_Band <- factor(new_data\$HFRS_Band, levels = levels(tertiary_revisions\$HFRS_Band)) new_data\$sex <- factor(new_data\$sex, levels = levels(tertiary_revisions\$sex)) new_data\$FinY <- factor(new_data\$FinY, levels = levels(tertiary_revisions\$FinY)) new_data\$infection <- factor(new_data\$infection, levels = levels(tertiary_revisions\$infection))</pre>
#Factors are consistent with model
levels(new_data\$HFRS_Band) levels(tertiary_revisions\$HFRS_Band)
levels(new_data\$sex) levels(tertiary_revisions\$sex)
levels(new_data\$FinY) levels(tertiary_revisions\$FinY)
levels(new_data\$ProvCode) levels(tertiary_revisions\$ProvCode)
levels(new_data\$infection) levels(tertiary_revisions\$infection)
<pre># Check levels of ProvCode in both datasets setdiff(levels(new_data\$ProvCode), levels(tertiary_revisions\$ProvCode)) # Levels in new_data but not in tertiary_revisions setdiff(levels(tertiary_revisions\$ProvCode), levels(new_data\$ProvCode)) # Levels in tertiary_revisions but not in new_data</pre>
new_data\$ProvCode <- droplevels(new_data\$ProvCode) # Check for missing values in factor variables sum(is.na(new_data\$ProvCode)) # Number of missing values in ProvCode
Ensure that ProvCode is a factor new_data\$ProvCode <- factor(new_data\$ProvCode, levels = levels(tertiary_revisions\$ProvCode))

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

# Now try th	he prediction again
predicted_p	probs <- predict(model_with_custom_splines, newdata = new_data, type =
"response"))
# Combine r	mean_unit_range and predicted_probs into a data frame
plot_data <·	- data.frame(DistanceMiles = DistanceMiles_range, predicted_prob =
predicted_p	probs)
#Calculate 9	95% confidence intervals
# Obtain pre	edicted values and standard errors for the new data
predictions	<- predict(model_with_custom_splines, newdata = new_data, type = "link",
se.fit = TRU	E)
# Calculate f	the confidence intervals for the log-odds scale (link scale)
# Use a 95%	6 confidence level (z-value = 1.96 for a 95% CI)
z_value <- 1	1.96
log_odds_lc	ower <- predictions\$fit - z_value * predictions\$se.fit
log_odds_u	pper <- predictions\$fit + z_value * predictions\$se.fit
# Convert th	ne log-odds confidence intervals to probabilities
# First, appl [,]	y the inverse link function (logistic function) to the log-odds
lower_prob	<- plogis(log_odds_lower)
upper_prob	o <- plogis(log_odds_upper)
<pre># Combine t plot_data <- DistanceM predicted_ ci_lower = ci_upper =)</pre>	the predicted probabilities and their confidence intervals into a data frame - data.frame(liles = new_data\$DistanceMiles, _prob = plogis(predictions\$fit), # Logistic transformation of the link lower_prob, upper_prob
# Combine r plot_data <-	mean_unit_range, predicted_probs, ci_lower, and ci_upper into plot_data - data.frame(DistanceMiles = DistanceMiles_range, predicted_prob = predicted_probs, ci_lower = boot_results\$ci_lower, ci_upper = boot_results\$ci_upper)
library(ggplo	ot2)
# Plot the sp	pline curve with confidence intervals
ggplot(plot_	_data, aes(x = DistanceMiles)) +

```
2
3
              geom line(aes(y = predicted prob), color = "blue", size = 1) +
4
              geom ribbon(aes(ymin = ci lower, ymax = ci upper), fill = "blue", alpha = 0.2) +
5
              labs(x = "Distance (Miles)", y = "Predicted Probability Readmission at 30 days") +
6
7
              theme minimal()
8
9
             library(dplyr)
10
11
             # Group by mean unit and calculate mean predicted prob and corresponding confidence
12
13
             intervals
14
             mean data <- plot data %>%
15
              group by(DistanceMiles) %>%
16
              summarise(
17
18
               mean predicted prob = mean(predicted prob, na.rm = TRUE),
19
               mean ci lower = mean(ci lower, na.rm = TRUE),
20
               mean ci upper = mean(ci upper, na.rm = TRUE)
21
              )
22
23
24
25
             # Define specific breaks (e.g., 25, 50, 75, ..., up to the maximum)
26
             breaks seq <- seq(0, max(mean data$DistanceMiles, na.rm = TRUE), by = 5)
27
28
29
             library(ggplot2)
30
             # Plot with specified increments on x-axis
31
             ggplot(mean_data, aes(x = DistanceMiles, y = mean_predicted_prob)) +
32
              geom point() + # Add points for mean predicted prob
33
              geom line() + # Connect points with a line
34
35
              geom_ribbon(aes(ymin = mean_ci_lower, ymax = mean_ci_upper), fill = "blue", alpha =
36
             0.2) + # Add ribbon for confidence intervals
37
              labs(x = "Travel Distance (Miles)", y = "Mean Predicted Probability for readmission at 30
38
             days", title = "Spline curve predicted probability of readmission at 30 days by patient travel
39
             distance") +
40
41
              scale x continuous(limits = c(0, max(mean data$DistanceMiles, na.rm = TRUE)), breaks =
42
             breaks seq) +
43
              theme minimal() +
44
              theme(
45
46
               axis.title.x = element text(size = 14), # Increase x-axis title font size
47
               axis.title.y = element text(size = 14), # Increase y-axis title font size
48
               axis.text.x = element text(size = 12), # Increase x-axis tick label font size
49
               axis.text.y = element text(size = 12), # Increase y-axis tick label font size
50
               plot.title = element text(size = 16, hjust = 0.5) # Increase plot title font size and center it
51
52
              )
53
54
55
             #Spline curve does appear to show the predicted probability of emergency readmission at
56
57
             30 days increases with travel distance but wide confidence intervals
58
59
             #Model Distance Miles and 30 day readmission with 3 knot splines
60
```

####First Imputation and descriptive stats####
#Use first imputed data for clinical and demographic characteristic summary
#complete_data is the first imputation
Count unique levels of ProvCode n_levels <- length(unique(complete_data\$ProvCode)) cat("Number of unique providers (ProvCode):", n_levels, "\n")
Count unique levels of sites n_levels <- length(unique(complete_data)) cat("Number of unique providers (ProvCode):", n_levels, "\n")
Count unique levels of ProvCode n_levels <- length(unique(tertiary_revisions\$ProvCode)) cat("Number of unique providers (ProvCode):", n_levels, "\n")
#38 unique providers
#Number of sites
Count unique levels of Sites but need to use original dataframe as sites not included in imputation analysis
#Find all those attending tertirary referral centre from original data tertiary_all <- subset(combined_data, MRC == 1)
#Find number of sites n_levels <- length(unique(tertiary_all\$Sitecode)) cat("Number of unique providers (Sites):", n_levels, "\n")
#187 sites
#Back to first imputation dataset. Calculate median number of miles straight line distance
summary(complete_data\$DistanceMiles)
#Median is 7.1 IQR is 3.9 to 12.7. Range 0 to 77.1 miles.

distance

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

#Driving distances

#Median 10.4 miles, IQR is 5.8 to 18.3 miles #Calculate median driving times summary(complete_data\$PeakDriveTime) #Median is 27 minutes IQR is 18.4 to 38.4. Maximum 104 minutes #Create travel time quintile variable quintiles <- quantile(complete_data\$DistanceMiles, probs = seq(0,1,0.2), na.rm=TRUE) complete data\$distancequintile <- cut(complete data\$DistanceMiles, breaks = quintiles, ..lowe labels = c("Q1", "Q2", "Q3", "Q4", "Q5"), include.lowest = TRUE) #Tabulate descriptive stats hist(tertiary_all\$Spell_Los) summary(tertiary_all\$Spell_Los) # Total number of revisions total_revisions <- nrow(complete_data)</pre> # Create a summary table summary_stats <- complete_data %>% group by(distancequintile) %>% summarise(# Count of observations Count = n(), **#** Distinct Providers Distinct Units = n distinct(ProvCode), Total Distinct Units = n distinct(complete data\$ProvCode), Distinct_Units_Percent = (Distinct_Units / Total_Distinct_Units) * 100, #Median distance Distance LowerQuartile = quantile(DistanceMiles, 0.25, na.rm = TRUE), Distance Median = median(DistanceMiles, na.rm = TRUE), Distance UpperQuartile = quantile(DistanceMiles, 0.75, na.rm = TRUE),

summary(complete data\$OffPeakDriveDistanceMiles)

2	
3	#Moon driving time
4	
5	DrivingTime_LowerQuartile = quantile(PeakDriveTime, 0.25, na.rm = TRUE),
6	DrivingTime_Median = median(PeakDriveTime, na.rm = TRUE),
7	DdrivingTime UpperQuartile = quantile(PeakDriveTime, 0.75, na.rm = TRUE),
8	o <u>-</u> i i i i i i i i i i
9	
10	
11	
12	# Age: Mean and standard deviation
13	Age Mean = mean(age of patient, na.rm = TRUE),
14	Age SD = sd(age of patient, na, rm = TRUF).
15	
16	
17	# Age: Mean ± SD (concatenated)
18	Age_Mean_SD = paste(round(mean(age_of_patient, na.rm = TRUE), 2), "±",
19	round(sd(age_of_patient, na.rm = TRUE), 2)),
20	
21	
22	# Condex frequency and a second as
23	# Gender: frequency and percentage
24	Female_Freq = sum(sex == "Female", na.rm = TRUE),
25	Female_Percent = sum(sex == "Female", na.rm = TRUE) / n() * 100,
26	Male Freg = sum(sex == "Male", na.rm = TRUE),
27	Male Percent = sum(sex == "Male" na rm = TRUE) / $n() * 100$
28	Male_refeent = suff(sex == male , numm = moly / n() = 100,
29	
21	# ASA: frequency and percentage for each level
37	HFRS_None_Freq = sum(HFRS_Band == "None", na.rm = TRUE),
32	HFRS_None_Percent = sum(HFRS_Band == "None", na.rm = TRUE) / n() * 100,
34	HFRS Mild Freg = sum(HFRS Band == "Mild", na.rm = TRUE),
35	HERS Mild Percent = sum(HERS Band == "Mild" na rm = TRUE) / $n() * 100$
36	HEPS Moderate Freq = sum(HEPS Band == "Moderate" na rm = TPLIE)
37	HERS_MODELALE_FIEL = SUIII(HERS_BAIL = MODELALE, HALL = HOE),
38	HFRS_Moderate_Percent = sum(HFRS_Band == "Moderate", na.rm = IRUE) / n() * 100,
39	HFRS_Severe_Freq = sum(HFRS_Band == "Severe", na.rm = TRUE),
40	HFRS_Severe_Percent = sum(HFRS_Band == "Severe", na.rm = TRUE) / n() * 100,
41	
42	
43	Historian
44	#Intection
45	
46	Infection_Freq = sum(infection == "1", na.rm = TRUE),
47	Infection Percent = sum(infection == "1", na.rm = TRUE) / n() * 100,
48	
49	
50	# Maan franker and a successfan as shows a frank 2000 to 2010
51	# Year: frequency and percentage for each year from 2009 to 2019
52	Year_2015_2016_Freq = sum(FinY == "2015/16", na.rm = TRUE),
55 54	Year_2015_2016_Percent = sum(FinY == "2015/16", na.rm = TRUE) / n() * 100,
55 55	Year 2016 2017 Freq = sum(FinY == "2016/17", na.rm = TRUE),
55	Year 2016 2017 Percent = sum(FinY == "2016/17", na rm = TRUF) / n() * 100
57	$V_{\text{par}} = 2017 - 2018 \text{ Freq} = \text{sum}(\text{EinV} = -2017/10^{\circ} \text{ na rm} = \text{TPLE})$
58	$\frac{1}{201} = \frac{1}{2010} = \frac{1}{100} = 1$
59	rear_201/_2018_Percent = $sum(FinY == 201//18^{\circ}, na.rm = IRUE) / n() * 100,$
60	Year_2018_2019_Freq = sum(FinY == "2018/19", na.rm = TRUE),

Year_2018_2019_Percent = sum(FinY == "2018/19", na.rm = TRUE) / n() * 100, Year_2019_2020_Freq = sum(FinY== "2019/20", na.rm = TRUE), Year_2019_2020_Percent = sum(FinY == "2019/20", na.rm = TRUE) / n() * 100,
Median Surgeon Volume: lower quartile, median, and upper quartile Surgeon_LowerQuartile = quantile(CV12mo, 0.25, na.rm = TRUE), Surgeon_Median = median(CV12mo, na.rm = TRUE), Surgeon_UpperQuartile = quantile(CV12mo, 0.75, na.rm = TRUE),
#Median hospital volume
Hospital_LowerQuartile = quantile(TV12mo, 0.25, na.rm = TRUE), Hospital_Median = median(TV12mo, na.rm = TRUE), Hospital_UpperQuartile = quantile(TV12mo, 0.75, na.rm = TRUE),
#Median IMD Score
IMD_LowerQuartile = quantile(IMD_score, 0.25, na.rm = TRUE), IMD_Median = median(IMD_score, na.rm = TRUE), IMD_UpperQuartile = quantile(IMD_score, 0.75, na.rm = TRUE),
)
Print the summary table print(summary_stats)
write.csv(summary_stats, "/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/Summary_stats.csv")
####Cluster Variable ####
<pre># Compute the mean outcome for each cluster library(dplyr) prov_means <- tertiary_revisions %>% group_by(ProvCode) %>% summarize(mean_outcome = mean(Read30days, na.rm = TRUE)) # Plot variability</pre>

2	
3	hoxplot(mean_outcome ~ ProvCode_data = prov_means_xlab = "ProvCode"_vlab = "Mean
4	
5	Outcome)
6	
7	# Summary statistics of variability
8	
0	summary(prov_means\$mean_outcome)
9 10	
10	#There is evidence of variability between providers
11	in there is evidence of valiability between providers
12	
13	
14	# Fit logistic regression on imputed datasets
15	m2 mi < with/tentient, revisions, almer/Deed20deve & DistanceMiles + IMD, seens +
16	m3.ml <- with(tertiary_revisions, gimer(kead30days ~ Distanceivilies + liviD_score +
17	HFRS_Band +
18	sex + age of patient + infection + TV12mo + CV12mo + FinY + (1
19	
20	Proveodej,
20	family = "binomial"))
21	
22	
23	
24	print(m3.mi)
25	
26	Hinduding DrayCode as a random affect was tasted but led to convergence issues likely due
27	#including Proveoue as a random effect was tested but led to convergence issues likely due
28	to numerical instability between providers so a decision was made to accept the fixed
29	effects model which may account for clustering at the provider level but is a limitation of
30	the study
31	
37	
22	
22	#Was travel distance strengly correlated with IMD score or ago?
34	#was travel distance strongly correlated with hvid_score of age:
35	
36	
37	
38	
39	#Next do a Spearman's rank correlation between travel distance and age, and then for
40	travel distance and IMD score
41	
42	
43	imp2\$MRC <- ifelse(imp2\$1V12mo > 49, 1, 0)
44	
15	tertiary revisions <- subset(imn2_MBC == 1)
45	$certary_revisions < subset(mp2, where = 1)$
40	
4/	
48	write csy/tertiary_revisions_"/Lisers/alexandermatthews//OneDrive - University of
49	Fustor (Alex Matthews MD / Devision Knoc Naturalis MD / Travel Times
50	Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times
51	Analysis_/tertiary_revisions.csv")
52	
53	
54	
55	tertiary_revisions <- as.mids(tertiary_revisions)
56	
57	tortions revisions for a patient a
57	ter tiar y_revisionspage_or_patient <-
00	as.numeric(as.character(tertiary_revisions\$age_of_patient))
59	
60	

tertiary revisions\$DistanceMiles <as.numeric(as.character(tertiary revisions\$DistanceMiles)) #Age and travel distance, Cannot pool the results based on the multiple imputations as cor test not compatible. Therefore stack all imputations together and calculate correlation # Scatterplot with linear regression line plot(tertiary revisions\$age of patient, tertiary revisions\$DistanceMiles, main = "Scatterplot of Age of Patient vs DistanceMiles", xlab = "Age of Patient", ylab = "DistanceMiles", pch = 19, col = "blue"# Add a linear trendline abline(Im(DistanceMiles ~ age of patient, data = tertiary revisions), col = "red", lwd = 2) # Calculate Spearman's rank correlation spearman test <- cor.test(tertiary revisions\$age of patient, tertiary revisions\$DistanceMiles, method = "spearman") # Extract rho and p-value rho <- round(spearman_test\$estimate, 2)</pre> p value <- spearman test\$p.value p_value_text <- ifelse(p_value < 0.05, "<0.05", paste0("=", round(p_value, 3)))</pre> # Add a legend with Spearman's rank correlation information legend("topright", legend = paste("Spearman's Rank Correlation:\n", "rho =", rho, ", p-value", p_value_text), col = c("blue", "red"), lty = c(NA, 1), pch = c(19, NA), lwd = c(NA, 2), bty = "n") #IMD score and travel distance # Scatterplot with trendline plot(tertiary revisions\$IMD score, tertiary revisions\$DistanceMiles, main = "Scatterplot of IMD score vs DistanceMiles", xlab = "IMD score", ylab = "DistanceMiles", pch = 19, col = "blue"# Add a linear trendline (for visualizing the general trend) abline(Im(DistanceMiles ~ IMD score, data = tertiary revisions), col = "red", lwd = 2) # Calculate Spearman's rank correlation spearman_test <- cor.test(tertiary_revisions\$IMD_score, tertiary_revisions\$DistanceMiles, method = "spearman")

2	
3	
4	# Extract rho and p-value
5	rho < round(snoorman_tost¢ostimato_2)
6	The control of the second se
/	p_value <- spearman_test\$p.value
8	p_value_text <- ifelse(p_value < 0.05, "<0.05", paste0("=", round(p_value, 3)))
9	
10	# Add a legend with Spearman's rank correlation information
12	legend("topright", legend = paste("Spearman's Rank Correlation:\n",
13	"rho -" rho " n-value" n value text)
14	rol = c("blue" "rod") [ty = c(NA 1) rob = c(10 NA) by d = c(NA 2) bty = "p")
15	cor = c(brue, red), red = c(nA, 1), pcr = c(19, nA), rwd = c(nA, 2), bry = rr)
16	
17	#Exposure 2 - OffPeakDriveDistanceMiles
18	
19	library("Ime4")
20	
21	# Eit logistic regression on imputed datasets include ProvCode in fixed effects to account for
22	# The logistic regression on impliced datasets include Proveode in fixed effects to account for
23	clustering
24	m3.mi <- with(tertiary_revisions, glm(Read30days ~ OffPeakDriveDistanceMiles +
25	IMD_score + HFRS_Band +
20	sex + age_of_patient + infection + TV12mo + CV12mo + FinY +
27	ProvCode.
20	family = "hinomial"))
30	
31	
32	
33	print(m3.mi)
34	
35	
36	# Pool results across imputed datasets
37	nooled results <- nool(m3 mi)
38	
39	
40	# Summarize pooled results with confidence intervals
41	summary_pooled <- summary(pooled_results, conf.int = TRUE)
42 43	
44	# Add Odds Ratios to the summary
45	summary pooledSOR <- exp(summary pooledSestimate)
46	summary pooled slower CI <- exp(summary pooled $(2,5\%)$)
47	summary_pooled\$Lipper_CL<_exp(summary_pooled\$`2.578')
48	summary_pooleu\$opper_cr <- exp(summary_pooleu\$ 37.5 %)
49	
50	# Display the final table with Odds Ratios and Confidence Intervals
51	print(summary_pooled)
52	
53	#check for evidence of multicollinearity?
54	, ,
55	library(car)
50 57	
58 58	
50	# Use the first imputed dataset for the VIF calculation
60	complete_data <- complete(tertiary_revisions, 1)

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

```
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

```
# Fit a logistic regression model on the complete dataset
vif model <- glm(Read30days ~ OffPeakDriveDistanceMiles + IMD score + HFRS Band +
          sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode,
         data = complete data, family = "binomial")
# Calculate VIF
vif values <- vif(vif model)
print(vif_values)
#No evidence of multi-collinearity
#Is there a non linear relationship?
# Custom function to add log-transformed variable and interaction term
add interaction <- function(data) {
 data$Log OffPeakDriveDistanceMiles <- log(data$OffPeakDriveDistanceMiles) # Add log-
transformed variable
 data$Interaction <- data$OffPeakDriveDistanceMiles *</pre>
data$Log_OffPeakDriveDistanceMiles # Add interaction term
 return(data)
}
# Extract the long-format data including the original data
tertiary_revisions_modified <- complete(tertiary_revisions, action = "long", include = TRUE)
# Apply the transformation to each imputed dataset
tertiary revisions modified <- do.call("rbind",
                     lapply(split(tertiary_revisions_modified,
tertiary revisions modified$.imp),
                         add interaction))
# Convert back to mids object
tertiary_revisions_modified <- as.mids(tertiary_revisions_modified)
# Fit the logistic regression model with the interaction term
model <- with(tertiary revisions modified, glm(Read30days ~ OffPeakDriveDistanceMiles +
Interaction,
                          family = binomial(link = "logit")))
# Pool the results
pooled results <- pool(model)
# Summarize pooled results
summary_pooled <- summary(pooled_results, conf.int = TRUE)</pre>
```

BMJ Open

# Extra box_ti	act the p-value for the interaction term idwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]
# Prin [.] print(l	t the p-value box_tidwell_p)
# p va	lue = 0.05. There is no evidence of non linearity
#Expo	osure 3 - PeakDriveTime
libran	/("Ime4")
library	
# Fit lo cluste	ogistic regression on imputed datasets include ProvCode in fixed effects to account for ring
m3.mi HFRS_	i <- with(tertiary_revisions, glm(Read30days ~ PeakDriveTime + IMD_score + _Band +
– ProvC	sex + age_of_patient + infection + TV12mo + CV12mo + FinY +
	family = "binomial"))
print(i	m3.mi)
# Pool poolee	l results across imputed datasets d_results <- pool(m3.mi)
# Sum summ	marize pooled results with confidence intervals hary_pooled <- summary(pooled_results, conf.int = TRUE)
# Add summ summ summ	Odds Ratios to the summary hary_pooled\$OR <- exp(summary_pooled\$estimate) hary_pooled\$Lower_CI <- exp(summary_pooled\$`2.5 %`) hary_pooled\$Upper_CI <- exp(summary_pooled\$`97.5 %`)
# Disp print(s	play the final table with Odds Ratios and Confidence Intervals summary_pooled)
#chec	k for evidence of multicollinearity?
library	y(car)
# Use compl	the first imputed dataset for the VIF calculation lete_data <- complete(tertiary_revisions, 1)
```
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

```
# Fit a logistic regression model on the complete dataset
vif model <- glm(Read30days ~ PeakDriveTime + IMD score + HFRS Band +
          sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode,
         data = complete data, family = "binomial")
# Calculate VIF
vif values <- vif(vif model)
print(vif_values)
#No evidence of multi-collinearity
#Is there a non linear relationship?
# Custom function to add log-transformed variable and interaction term
add interaction <- function(data) {
 data$Log PeakDriveTime <- log(data$PeakDriveTime) # Add log-transformed variable
 data$Interaction <- data$PeakDriveTime * data$Log PeakDriveTime # Add interaction
term
 return(data)
}
# Extract the long-format data including the original data
tertiary_revisions_modified <- complete(tertiary_revisions, action = "long", include = TRUE)
# Apply the transformation to each imputed dataset
tertiary revisions modified <- do.call("rbind",
                     lapply(split(tertiary_revisions_modified,
tertiary revisions modified$.imp),
                         add interaction))
# Convert back to mids object
tertiary_revisions_modified <- as.mids(tertiary_revisions_modified)
# Fit the logistic regression model with the interaction term
model <- with(tertiary_revisions_modified, glm(Read30days ~ PeakDriveTime + Interaction,
                          family = binomial(link = "logit")))
# Pool the results
pooled results <- pool(model)
# Summarize pooled results
summary_pooled <- summary(pooled_results, conf.int = TRUE)</pre>
```

2	
3	
4	# Extract the p-value for the interaction term
5	box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]
6	
7	# Print the p-value
8	print/box tidwell n)
9	print(box_tidweil_p)
10	
11	# p value = 0.13 not evidence of non linearity
12	
13	
14	
15	
16	
17	
18	####Secondary Outcome mortality 90 days ####
19	
20	
21	#Evnesure 1 Distance Miles
22	#Exposure 1 - Distance Miles
23	
24	library("lme4")
25	
26	# Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for
27	clustoring
28	
29	m3.ml <- with(tertiary_revisions, glm(Mort90days ~ DistanceMiles + IMD_score +
30	HFRS_Band +
31	sex + age_of_patient + infection + TV12mo + CV12mo + FinY +
2∠ 22	ProvCode.
34	family = "hinomial"))
35	
36	
37	La
38	print(m3.mi)
39	
40	
41	# Pool results across imputed datasets
42	noolod rosults <- nool(m3 mi)
43	pooled_results <- pool(ins.ini)
44	
45	# Summarize pooled results with confidence intervals
46	summary_pooled <- summary(pooled_results, conf.int = TRUE)
47	
48	# Add Odds Ratios to the summary
49	r_{r} rad odds hallos to the summary pooled (estimate)
50	summary_pooleusork <- exp(summary_pooleusestimate)
51	summary_pooled\$Lower_Cl <- exp(summary_pooled\$ 2.5 %)
52	summary_pooled\$Upper_CI <- exp(summary_pooled\$`97.5 %`)
53	
54	# Display the final table with Odds Ratios and Confidence Intervals
55 56	nrint(summary_nooled)
50 57	
57 50	
50	#check for evidence of multicollinearity?
59	
00	

```
library(car)
# Use the first imputed dataset for the VIF calculation
complete_data <- complete(tertiary_revisions, 1)
# Fit a logistic regression model on the complete dataset
vif model <- glm(Mort90days ~ DistanceMiles + IMD score + HFRS Band +
          sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode,
         data = complete_data, family = "binomial")
# Calculate VIF
vif values <- vif(vif model)
print(vif_values)
#No evidence of multi-collinearity
#Is there evidence of non linearity?
library(mice)
tertiary_revisions <- as.mids(tertiary_revisions)
# Custom function to add log-transformed variable and interaction term
add interaction <- function(data) {
 data$Log DistanceMiles <- log(data$DistanceMiles) # Add log-transformed variable
 data$Interaction <- data$DistanceMiles * data$Log DistanceMiles # Add interaction term
 return(data)
}
# Extract the long-format data including the original data
tertiary revisions modified <- complete(tertiary revisions, action = "long", include = TRUE)
# Apply the transformation to each imputed dataset
tertiary revisions modified <- do.call("rbind",
                     lapply(split(tertiary_revisions_modified,
tertiary revisions modified$.imp),
                         add interaction))
# Convert back to mids object
tertiary revisions modified <- as.mids(tertiary revisions modified)
# Fit the logistic regression model with the interaction term
model <- with(tertiary revisions modified, glm(Mort90days ~ DistanceMiles + Interaction,
                          family = binomial(link = "logit")))
```

Pool the results pooled_results <- pool(model)
Summarize pooled results summary_pooled <- summary(pooled_results, conf.int = TRUE)
Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]
Print the p-value print(box_tidwell_p)
P value 0.95
#Exposure 2 - OffPeakDriveDistanceMiles
library("lme4")
Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for clustering m3.mi <- with(tertiary revisions, glm(Mort90days ~ OffPeakDriveDistanceMiles +
IMD_score + HFRS_Band + sex + age_of_patient + infection + TV12mo + CV12mo + FinY +
ProvCode, family = "binomial"))
print(m3.mi)
Pool results across imputed datasets pooled_results <- pool(m3.mi)
Summarize pooled results with confidence intervals summary_pooled <- summary(pooled_results, conf.int = TRUE)
Add Odds Ratios to the summary summary_pooled\$OR <- exp(summary_pooled\$estimate) summary_pooled\$Lower_CI <- exp(summary_pooled\$`2.5 %`) summary_pooled\$Upper_CI <- exp(summary_pooled\$`97.5 %`)
Display the final table with Odds Ratios and Confidence Intervals print(summary_pooled)

1 2	
3	
4	#check for evidence of multicollinearity?
5	
7	librarv(car)
8	
9	# Use the first imputed dataset for the VIE calculation
10	complete data <- complete(tertiary revisions 1)
11	
13	# Fit a logistic regression model on the complete dataset
14	vif model <- glm(Read30days ~ OffPeakDriveDistanceMiles + IMD_score + HERS_Band +
15	sex + age of natient + infection + TV12mo + CV12mo + FinY + ProvCode
16	data = complete_data_family = "hinomial")
17	
19	# Calculate VIE
20	vif values <- vif(vif model)
21	nrint(vif_values)
22	
23	
25	#No evidence of multi-collinearity
26	
27	#Is there evidence of non linearity?
28 29	institute evidence of non-incurry.
30	tertiary revisions <- as mids(tertiary revisions)
31	
32	# Custom function to add log-transformed variable and interaction term
33	add interaction <- function(data) {
35	data\$1.0g_OffPeakDriveDistanceMiles <= log(data\$OffPeakDriveDistanceMiles) # Add log_
36	transformed variable
37	dataSinteraction <- dataSOffPeakDriveDistanceMiles *
38	data\$l.og_OffPeakDriveDistanceMiles_#Add_interaction_term
39 40	return(data)
41	
42	
43	# Extract the long-format data including the original data
44 45	tertiary revisions modified <- complete/tertiary revisions action = "long" include = TRUE)
45 46	
47	# Apply the transformation to each imputed dataset
48	π Apply the transformation to each imputed dataset
49	lannly/snlit/tertiary revisions modified
50 51	tortiary rovisions modified\$ imp)
52	add interaction)
53	
54	# Convert back to mids object
55	π convert back to milds object tertiary revisions modified z_{-} as mids/tertiary revisions modified)
56 57	
58	# Fit the legistic regression model with the interaction term
59	
60	

model <- with(tertiary_revisions_modified, glm(Mort90days ~ OffPeakDriveDistanceMiles + Interaction.
family = binomial(link = "logit")))
Pool the results pooled_results <- pool(model)
Summarize pooled results summary_pooled <- summary(pooled_results, conf.int = TRUE)
Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]
Print the p-value print(box_tidwell_p)
#0.989
#Exposure 3 - PeakDriveTime
library("lme4")
Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for clustering
m3.mi <- with(tertiary_revisions, glm(Mort90days ~ PeakDriveTime + IMD_score + HFRS_Band +
sex + age of patient + infection + TV12mo + CV12mo + FinY +
ProvCode,
family = "binomial"))
print(m3.mi)
Pool results across imputed datasets pooled_results <- pool(m3.mi)
Summarize pooled results with confidence intervals summary_pooled <- summary(pooled_results, conf.int = TRUE)
Add Odds Ratios to the summary summary_pooled\$OR <- exp(summary_pooled\$estimate) summary_pooled\$Lower_CI <- exp(summary_pooled\$`2.5 %`)

```
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

summary_pooled\$Upper_Cl <- exp(summary_pooled\$`97.5 %`)</pre>

Display the final table with Odds Ratios and Confidence Intervals
print(summary_pooled)

#check for evidence of multicollinearity?

library(car)

Use the first imputed dataset for the VIF calculation complete_data <- complete(tertiary_revisions, 1)</pre>

Calculate VIF
vif_values <- vif(vif_model)
print(vif_values)</pre>

#No evidence of multi-collinearity

```
#Is there evidence of non linearity?
```

```
# Custom function to add log-transformed variable and interaction term
add_interaction <- function(data) {
    data$Log_PeakDriveTime <- log(data$PeakDriveTime) # Add log-transformed variable
    data$Interaction <- data$PeakDriveTime * data$Log_PeakDriveTime # Add interaction</pre>
```

term

return(data)

}

Extract the long-format data including the original data tertiary_revisions_modified <- complete(tertiary_revisions, action = "long", include = TRUE)</pre>

```
# Convert back to mids object
tertiary_revisions_modified <- as.mids(tertiary_revisions_modified)</pre>
```

Fit the logistic regression model with the interaction term

model <- with(tertiary_revisions_modified, glm(Mort90days ~ PeakDriveTime + Interaction, family = binomial(link = "logit")))
Pool the results pooled_results <- pool(model)
Summarize pooled results summary_pooled <- summary(pooled_results, conf.int = TRUE)
Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]
Print the p-value print(box_tidwell_p)
P avlue 0.78
####Secondary outcome prolonged LOS ####
tertiary_revisions <- complete(tertiary_revisions, "long", inc = TRUE)
tertiary_revisions\$Long_Los <- ifelse(tertiary_revisions\$Spell_Los > 5, 1, 0)
tertiary_revisions <- as.mids(tertiary_revisions)
#Exposure 1 - Distance Miles
library("Ime4")
Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for clustering m3.mi <- with(tertiary revisions, glm(Long Los ~ DistanceMiles + IMD score + HFRS Band +
sex + age_of_patient + infection + TV12mo + CV12mo + FinY +
family = "binomial"))
print(m3.mi)
Pool results across imputed datasets pooled_results <- pool(m3.mi)
Summarize pooled results with confidence intervals summary_pooled <- summary(pooled_results, conf.int = TRUE)

```
# Add Odds Ratios to the summary
summary pooled$OR <- exp(summary pooled$estimate)
summary pooled$Lower_Cl <- exp(summary_pooled$`2.5 %`)</pre>
summary_pooled$Upper_Cl <- exp(summary_pooled$`97.5 %`)</pre>
# Display the final table with Odds Ratios and Confidence Intervals
print(summary_pooled)
#check for evidence of multicollinearity?
library(car)
# Use the first imputed dataset for the VIF calculation
complete data <- complete(tertiary revisions, 1)
# Fit a logistic regression model on the complete dataset
vif model <- glm(Long Los ~ DistanceMiles + IMD score + HFRS Band +
          sex + age_of_patient + infection + TV12mo + CV12mo + FinY + ProvCode,
         data = complete data, family = "binomial")
# Calculate VIF
vif_values <- vif(vif_model)</pre>
print(vif_values)
#No evidence of multi-collinearity
#Is there evidence of non linearity?
# Custom function to add log-transformed variable and interaction term
add interaction <- function(data) {
 data$Log DistanceMiles <- log(data$DistanceMiles) # Add log-transformed variable
 data$Interaction <- data$DistanceMiles * data$Log DistanceMiles # Add interaction term
 return(data)
}
# Extract the long-format data including the original data
tertiary revisions modified <- complete(tertiary revisions, action = "long", include = TRUE)
# Apply the transformation to each imputed dataset
tertiary revisions modified <- do.call("rbind",
                     lapply(split(tertiary_revisions_modified,
tertiary revisions modified$.imp),
                         add interaction))
# Convert back to mids object
tertiary revisions modified <- as.mids(tertiary revisions modified)
```

BMJ Open

Fit the logistic regression model with the interaction term

model <- with(tertiary_revisions_modified, glm(Long_Los ~ DistanceMiles + Interaction, family = binomial(link = "logit"))) # Pool the results pooled_results <- pool(model)</pre> # Summarize pooled results summary_pooled <- summary(pooled_results, conf.int = TRUE)</pre> # Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"] # Print the p-value print(box_tidwell_p) #P value 0.002 Non linear # Load the required library library(splines) #AIC of non spline model model <- glm(Long_Los ~ DistanceMiles, data = tertiary_revisions, family = binomial) summary(model) #AIC 52853 # Define a function to fit and evaluate spline models with knots based on centiles evaluate centile splines <- function(centiles, data) { # Calculate knots based on the specified centiles knots <- quantile(data\$DistanceMiles, probs = centiles, na.rm = TRUE)</pre> # Fit a logistic regression model with natural splines using the calculated knots model spline <- $glm(Long Los \sim ns(DistanceMiles, knots = knots)$, family = binomial(link = "logit"), data = data) # Summarize the model

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Summarize the model summary_model <- summary(model_spline)

Extract p-values for the spline terms
p_values <- summary_model\$coefficients[-1, "Pr(>|z|)"] # Exclude the intercept

Print the results

3	
4 5	
6	
7	
8	
9 10	
10	
12	
13	
14	
15	
17	
18	
19	
20	
22	
23	
24	
25 26	
27	
28	
29	
31	
32	
33	
34	
36	
37	
38	
39 40	
40 41	
42	
43	
44	
45 46	
47	
48	
49 50	
50	
52	
53	
54 55	
55 56	
57	
58	
59 60	
00	

cat("\nResults for centiles", centiles, ":\n") print(p_values)
<pre># Return the model and calculated knots for further inspection if needed return(list(model = model_spline, p_values = p_values, knots = knots)) }</pre>
Example centile configurations for 3, 4, and 5 knots centiles_3_knots <- c(0.05, 0.50, 0.95) # 5th, 50th, and 95th percentiles centiles_4_knots <- c(0.05, 0.35, 0.65, 0.95) # Custom centiles for 4 knots centiles_5_knots <- c(0.05, 0.25, 0.50, 0.75, 0.95) # 5 knots centiles
Evaluate models with centile-based knots using your dataset results_3_knots <- evaluate_centile_splines(centiles = centiles_3_knots, data = tertiary_revisions)
results_4_knots <- evaluate_centile_splines(centiles = centiles_4_knots, data =
tertiary_revisions) results_5_knots <- evaluate_centile_splines(centiles = centiles_5_knots, data = tertiary_revisions)
<pre># Compare models with centile-based knots cat("\nComparing models with different centile-based knots:\n") anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq")</pre>
<pre># Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots)</pre>
#52769, model with four knots best fit and improved fit from original linear model
#Run spline model with adjusted data excluding missing data library(splines)
<pre># For example, let's say you want 3 knots at specific percentiles knots <- quantile(tertiary_revisions\$DistanceMiles, probs = c(0.05, 0.35, 0.65, 0.95), na.rm = TRUE) print(knots)</pre>
spline_terms <- ns(tertiary_revisions\$DistanceMiles, knots = knots)

2	
3	model with custom splines <- glm(Long Los ~ ns(DistanceMiles, knots = knots) +
4	HERS Band + IMD score +
5	sor Lago of national TV/12mo LCV/12mo LEinV L DrovCodo
6	sex + age_or_patient + infection + 1v12mo + Cv12mo + Finv + ProvCode,
/	family = "binomial", data = tertiary_revisions)
8	
9	
10	summary(model with custom splines)
11	
12	#Concrete a sequence of mean unit values for predicting
14	#Generate a sequence of mean unit values for predicting
15	
16	DistanceMiles_range <- seq(min(tertiary_revisions\$DistanceMiles),
17	max(tertiary_revisions\$DistanceMiles), length.out = 100)
18	
19	new data <- expand grid(
20	Distance/Ailes - Distance/Ailes range
21	Distanceivilies = Distanceivilies_range,
22	sex = levels(tertiary_revisions\$sex), # Ensure it takes all factor levels
23	age_of_patient = mean(tertiary_revisions\$age_of_patient, na.rm = TRUE),
24	HFRS Band = levels(tertiary revisions\$HFRS Band), # Ensuring correct factor levels
25	IMD score = mean(tertiary revisionsSIMD score, $na, rm = TRUF$).
26	FinV = lovels(tertiany_revisions\$EinV) # Ensuring correct factor lovels
27	C) (12ma - magar (tertiany revisions) (11), # Linsuning correct factor levels
28	CV12mo = mean(tertiary_revisions\$CV12mo, na.rm = TRUE),
29	TV12mo = mean(tertiary_revisions\$TV12mo, na.rm = TRUE),
30	ProvCode = levels(tertiary_revisions\$ProvCode), # Ensuring correct factor levels
31	infection = levels(tertiary revisions\$infection) # Ensuring correct factor levels
32	
33	
34 25	
20	# Create a new dataset with a range of distances and miles and all other predictor variables
30	new_data <- expand.grid(DistanceMiles = DistanceMiles_range,
38	sex = unique(tertiary_revisions\$sex),
39	age of patient = mean(tertiary revisions\$age of patient),
40	HERS Band = unique(tertiary revisionsSHERS Band).
41	IMD_score = mean(tertiany_revisions\$IMD_score)
42	$\frac{1}{2} = \frac{1}{2} = \frac{1}$
43	Finy = unique(tertiary_revisions\$Finy),
44	CV12mo = mean(tertiary_revisions\$CV12mo),
45	TV12mo = mean(tertiary_revisions\$TV12mo),
46	infection = unique(tertiary revisions\$infection))
47	
48	
49	WAlter the lands of Dec. Code to see the terror to be the terror date.
50	# Align the levels of ProvCode in new_data to match the training data
51	new_data\$ProvCode <- factor(new_data\$ProvCode, levels =
52	levels(tertiary_revisions\$ProvCode))
53	
54	# Align the levels of all relevant categorical variables
55	new data\$HERS Band <- factor/new data\$HERS Band lovals -
50	new_uataşını ns_banu <- iactor(new_uataşınrns_banu, ievels -
5/ 59	ieveis(tertiary_revisions\$HFKS_Band))
20 50	new_data\$sex <- factor(new_data\$sex, levels = levels(tertiary_revisions\$sex))
59	new_data\$FinY <- factor(new_data\$FinY, levels = levels(tertiary_revisions\$FinY))
00	

2	
3	
4	
5	
ć	
0	
7	
8	
9	
10	
10	
11	
12	
13	
14	
14	
15	
16	
17	
10	
18	
19	
20	
21	
27	
22	
23	
24	
25	
25	
20	
27	
28	
29	
20	
30	
31	
32	
33	
24	
34	
35	
36	
37	
20	
38	
39	
40	
11	
40	
42	
43	
44	
15	
45	
46	
47	
48	
⊿0	
+7	
50	
51	
52	
52	
55	
54	
55	
56	
55	
5/	
58	
59	
60	

1

new_data\$infection <- factor(new_data\$infection, levels =
levels(tertiary_revisions\$infection))</pre>

#Factors are consistent with model

levels(new_data\$HFRS_Band) levels(tertiary_revisions\$HFRS_Band)

levels(new_data\$sex)
levels(tertiary_revisions\$sex)

levels(new_data\$FinY)
levels(tertiary_revisions\$FinY)

levels(new_data\$ProvCode) levels(tertiary_revisions\$ProvCode)

levels(new_data\$infection) levels(tertiary_revisions\$infection)

Check levels of ProvCode in both datasets
setdiff(levels(new_data\$ProvCode), levels(tertiary_revisions\$ProvCode)) # Levels in
new_data but not in tertiary_revisions
setdiff(levels(tertiary_revisions\$ProvCode), levels(new_data\$ProvCode)) # Levels in
tertiary_revisions but not in new_data

new_data\$ProvCode <- droplevels(new_data\$ProvCode)
Check for missing values in factor variables
sum(is.na(new_data\$ProvCode)) # Number of missing values in ProvCode</pre>

Ensure that ProvCode is a factor new_data\$ProvCode <- factor(new_data\$ProvCode, levels = levels(tertiary_revisions\$ProvCode))

Now try the prediction again
predicted_probs <- predict(model_with_custom_splines, newdata = new_data, type =
"response")</pre>

Combine mean_unit_range and predicted_probs into a data frame plot_data <- data.frame(DistanceMiles = DistanceMiles_range, predicted_prob = predicted_probs)

#Calculate 95% confidence intervals

4

5

6 7 8

9

10

11

12 13

14

15

16

17 18

19

20 21

22

23

24 25

26

27

28 29

36

37

38

39

40 41

42

43 44

45 46 47

48

49

50

51 52

53

54

55

56 57

```
# Obtain predicted values and standard errors for the new data
predictions <- predict(model with custom splines, newdata = new data, type = "link",
se.fit = TRUE)
# Calculate the confidence intervals for the log-odds scale (link scale)
# Use a 95% confidence level (z-value = 1.96 for a 95% CI)
z value <- 1.96
log odds lower <- predictions$fit - z value * predictions$se.fit
log_odds_upper <- predictions$fit + z_value * predictions$se.fit
# Convert the log-odds confidence intervals to probabilities
# First, apply the inverse link function (logistic function) to the log-odds
lower_prob <- plogis(log_odds_lower)</pre>
upper prob <- plogis(log odds upper)
# Combine the predicted probabilities and their confidence intervals into a data frame
plot data <- data.frame(
 DistanceMiles = new_data$DistanceMiles,
 predicted_prob = plogis(predictions$fit), # Logistic transformation of the link
 ci lower = lower prob,
 ci upper = upper prob
)
library(ggplot2)
# Plot the spline curve with confidence intervals
ggplot(plot data, aes(x = DistanceMiles)) +
 geom line(aes(y = predicted prob), color = "blue", size = 1) +
 geom_ribbon(aes(ymin = ci_lower, ymax = ci_upper), fill = "blue", alpha = 0.2) +
 labs(x = "Distance (Miles)", y = "Predicted Probability Readmission at 30 days") +
 theme minimal()
library(dplyr)
# Group by mean unit and calculate mean predicted prob and corresponding confidence
intervals
mean data <- plot data %>%
 group by(DistanceMiles) %>%
 summarise(
  mean_predicted_prob = mean(predicted_prob, na.rm = TRUE),
  mean ci lower = mean(ci lower, na.rm = TRUE),
  mean ci upper = mean(ci upper, na.rm = TRUE)
 )
```

BMJ Open

```
# Define specific breaks (e.g., 25, 50, 75, ..., up to the maximum)
breaks seq <- seq(0, max(mean data$DistanceMiles, na.rm = TRUE), by = 5)
library(ggplot2)
# Plot with specified increments on x-axis
ggplot(mean data, aes(x = DistanceMiles, y = mean predicted prob)) +
 geom point() + # Add points for mean predicted prob
 geom line() + # Connect points with a line
 geom_ribbon(aes(ymin = mean_ci_lower, ymax = mean_ci_upper), fill = "blue", alpha =
0.2) + # Add ribbon for confidence intervals
 labs(x = "Travel Distance (Miles)", y = "Mean Predicted Probability for Prolonged LOS", title
= "Spline curve predicted probability of prolonged LOS by patient travel distance") +
 scale x continuous(limits = c(0, max(mean data$DistanceMiles, na.rm = TRUE)), breaks =
breaks seq) +
 theme minimal() +
 theme(
  axis.title.x = element text(size = 14), # Increase x-axis title font size
  axis.title.y = element_text(size = 14), # Increase y-axis title font size
  axis.text.x = element text(size = 12), # Increase x-axis tick label font size
  axis.text.y = element_text(size = 12), # Increase y-axis tick label font size
  plot.title = element text(size = 16, hjust = 0.5) # Increase plot title font size and center it
 )
                                         reliew
#Exposure 2 - OffPeakDriveDistanceMiles
library("Ime4")
# Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for
clustering
m3.mi <- with(tertiary revisions, glm(Long Los ~ OffPeakDriveDistanceMiles + IMD score +
HFRS Band +
                      sex + age of patient + infection + TV12mo + CV12mo + FinY +
ProvCode,
                     family = "binomial"))
print(m3.mi)
# Pool results across imputed datasets
pooled results <- pool(m3.mi)
```

Summarize pooled results with confidence intervals summary_pooled <- summary(pooled_results, conf.int = TRUE)
Add Odds Ratios to the summary summary_pooled\$OR <- exp(summary_pooled\$estimate) summary_pooled\$Lower_Cl <- exp(summary_pooled\$`2.5 %`) summary_pooled\$Upper_Cl <- exp(summary_pooled\$`97.5 %`)
Display the final table with Odds Ratios and Confidence Intervals print(summary_pooled)
#check for evidence of multicollinearity?
library(car)
Use the first imputed dataset for the VIF calculation complete_data <- complete(tertiary_revisions, 1)
<pre># Fit a logistic regression model on the complete dataset vif_model <- glm(Read30days ~ DistanceMiles + IMD_score + HFRS_Band +</pre>
<pre># Calculate VIF vif_values <- vif(vif_model) print(vif_values)</pre>
#No evidence of multi-collinearity
#Is there evidence of non linearity?
<pre># Custom function to add log-transformed variable and interaction term add_interaction <- function(data) { data\$Log_OffPeakDriveDistanceMiles <- log(data\$OffPeakDriveDistanceMiles) # Add log- transformed variable data\$Interaction <- data\$OffPeakDriveDistanceMiles * data\$Log_OffPeakDriveDistanceMiles # Add interaction term return(data) }</pre>
Extract the long-format data including the original data tertiary_revisions, action = "long", include = TRUE)
Apply the transformation to each imputed dataset tertiary_revisions_modified <- do.call("rbind",

BMJ Open

lapply(split(tertiary_revisions_modified, tertiary revisions modified\$.imp), add interaction)) # Convert back to mids object tertiary revisions modified <- as.mids(tertiary revisions modified) # Fit the logistic regression model with the interaction term model <- with(tertiary_revisions_modified, glm(Long_Los ~ OffPeakDriveDistanceMiles + Interaction, family = binomial(link = "logit"))) # Pool the results pooled_results <- pool(model)</pre> # Summarize pooled results summary pooled <- summary(pooled results, conf.int = TRUE) # Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"] # Print the p-value elie print(box_tidwell_p) #0.003 #AIC of non spline model model <- glm(Long Los ~ OffPeakDriveDistanceMiles, data = tertiary revisions, family = binomial) summary(model) #AIC 52853 # Define a function to fit and evaluate spline models with knots based on centiles evaluate centile splines <- function(centiles, data) { # Calculate knots based on the specified centiles knots <- quantile(data\$OffPeakDriveDistanceMiles, probs = centiles, na.rm = TRUE) # Fit a logistic regression model with natural splines using the calculated knots model spline <- glm(Long Los ~ ns(OffPeakDriveDistanceMiles, knots = knots), family = binomial(link = "logit"), data = data) # Summarize the model summary model <- summary(model spline)

<pre># Extract p-values for the spline terms p_values <- summary_model\$coefficients[-1, "Pr(> z)"] # Exclude the in # Print the results cat("\nResults for centiles", centiles, ":\n") print(p_values) # Return the model and calculated knots for further inspection if needed return(list(model = model_spline, p_values = p_values, knots = knots)) # Example centile configurations for 3, 4, and 5 knots</pre>	ntercept
<pre>4 p_values is under p to doe op into to the opinite terms 5 p_values <- summary_model\$coefficients[-1, "Pr(> z)"] # Exclude the in 6 7 # Print the results 8 cat("\nResults for centiles", centiles, ":\n") 9 print(p_values) 10 11 12 # Return the model and calculated knots for further inspection if needed 13 return(list(model = model_spline, p_values = p_values, knots = knots)) 14 15 16 17 # Example centile configurations for 3, 4, and 5 knots</pre>	ntercept
<pre>5</pre>	l
<pre>6 7 # Print the results 8 cat("\nResults for centiles", centiles, ":\n") 9 print(p_values) 10 12 # Return the model and calculated knots for further inspection if needed 13 return(list(model = model_spline, p_values = p_values, knots = knots)) 14 } 15 16 17 # Example centile configurations for 3, 4, and 5 knots</pre>	I
<pre>7 # Print the results 8 cat("\nResults for centiles", centiles, ":\n") 9 print(p_values) 10 # Return the model and calculated knots for further inspection if needed 13 return(list(model = model_spline, p_values = p_values, knots = knots)) 14 } 15 # Example centile configurations for 3, 4, and 5 knots</pre>	
<pre>8 cat("\nResults for centiles", centiles, ":\n") 9 print(p_values) 10 12 # Return the model and calculated knots for further inspection if needed 13 return(list(model = model_spline, p_values = p_values, knots = knots)) 14 } 15 16 17 # Example centile configurations for 3, 4, and 5 knots</pre>	Ι
9 print(p_values) 10 print(p_values) 11 # Return the model and calculated knots for further inspection if needed 13 return(list(model = model_spline, p_values = p_values, knots = knots)) 14 } 15 16 17 # Example centile configurations for 3, 4, and 5 knots	Ι
<pre>10 print(p_values) 11 12 # Return the model and calculated knots for further inspection if needed 13 return(list(model = model_spline, p_values = p_values, knots = knots)) 14 } 15 16 17 # Example centile configurations for 3, 4, and 5 knots</pre>	1
<pre>11 12 # Return the model and calculated knots for further inspection if needed 13 return(list(model = model_spline, p_values = p_values, knots = knots)) 14 15 16 17 # Example centile configurations for 3, 4, and 5 knots</pre>	I
<pre>12 # Return the model and calculated knots for further inspection if needed 13 return(list(model = model_spline, p_values = p_values, knots = knots)) 14 } 15 16 17 # Example centile configurations for 3, 4, and 5 knots</pre>	l
<pre>return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots = knots)) return(list(model = model_spline, p_values = p_values, knots)) return(list(model = model_spline, p_values, knots))) return(list(model = model_spline, p_val</pre>	
 14 } 15 16 17 # Example centile configurations for 3, 4, and 5 knots 	
 15 16 17 # Example centile configurations for 3, 4, and 5 knots 	
 16 17 # Example centile configurations for 3, 4, and 5 knots 	
H Example centile configurations for 3, 4, and 5 knots	
1,	
18 contilos 2 knots $< < (0.05, 0.50, 0.95)$ # 5th 50th and 95th parcontilos	
19 centiles_4_knots <- c(0.05, 0.35, 0.65, 0.95) # Custom centiles for 4 knots	
centiles 5 knots <- c(0.05, 0.25, 0.50, 0.75, 0.95) # 5 knots centiles	
22 martine to the second line of the second line to using your detect	
23 # Evaluate models with centile-based knots using your dataset	
<pre>24 results_3_knots <- evaluate_centile_splines(centiles = centiles_3_knots, d</pre>	lata =
²⁵ tertiary revisions)	
$\frac{26}{26}$ results 4 knots <- evaluate centile splines(centiles = centiles 4 knots d	lata =
27 results_4_knots < evaluate_centile_splines(centiles = centiles_4_knots, a	
28 tertiary_revisions)	
29 results_5_knots <- evaluate_centile_splines(centiles = centiles_5_knots, d	lata =
³⁰ tertiary revisions)	
31	
32	
33 # Compare models with centile-based knots	
34 cat("\nComparing models with different centile-based knots:\n")	
35 anova(results 3 knots\$model, results 4 knots\$model, results 5 knots\$	model. test =
36 "Chica")	
37 Chisq)	
38	
39 # Print the calculated knot locations for each model	
40 cat("\nKnot locations for 3 knots:\n")	
41 print(results 2 knots(knots)	
42 print(results_3_knots)	
cat("\nKnot locations for 4 knots:\n")	
print(results 4 knots\$knots)	
45 cat/"\nKnot locations for 5 knots:\n")	
45 cat((minor locations for 5 khots, (if)	
⁴⁰ print(results_5_knots)	
4/	
48 #52718, model with four knots best fit and significant spline terms	
51 #Run spline model with adjusted data excluding missing data	
⁵² library(splines)	
⁵³ # For example, let's say you want 3 knots at specific percentiles	
54 knots < quantila/tartiany rovisions\$OffDaakDrivaDistanceMilas probs = 5	
55 knows \sim quantile (tertially_revisions; γ) reak DriveDistance villes, props = 0	.כס.ט, ככ.ט, כט.ט,
56 0.95), na.rm = TRUE)	
57 print(knots)	
58	
59 colling terms a politeritier revisions (Office Units Distance Miles Local	(moto)
60 spline_terms <- instructions_onreakDriveDistanceivilles, knots =	KHULSJ

```
model with custom splines <- glm(Long Los ~ ns(OffPeakDriveDistanceMiles, knots =
knots) + HFRS Band + IMD score +
                   sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode,
                  family = "binomial", data = tertiary_revisions)
summary(model_with_custom_splines)
#Generate a sequence of mean unit values for predicting
DistanceMiles_range <- seq(min(tertiary_revisions$OffPeakDriveDistanceMiles),
max(tertiary revisions$OffPeakDriveDistanceMiles), length.out = 100)
new data <- expand.grid(
 OffPeakDriveDistanceMiles = DistanceMiles range,
 sex = levels(tertiary revisions$sex), # Ensure it takes all factor levels
 age of patient = mean(tertiary revisions$age of patient, na.rm = TRUE),
 HFRS Band = levels(tertiary revisions$HFRS Band), # Ensuring correct factor levels
 IMD_score = mean(tertiary_revisions$IMD_score, na.rm = TRUE),
 FinY = levels(tertiary revisions$FinY), # Ensuring correct factor levels
 CV12mo = mean(tertiary_revisions$CV12mo, na.rm = TRUE),
 TV12mo = mean(tertiary revisions$TV12mo, na.rm = TRUE),
 ProvCode = levels(tertiary revisions$ProvCode), # Ensuring correct factor levels
 infection = levels(tertiary_revisions$infection) # Ensuring correct factor levels
)
# Create a new dataset with a range of distances and miles and all other predictor variables
new_data <- expand.grid(DistanceMiles = DistanceMiles_range,
             sex = unique(tertiary revisions$sex),
             age of patient = mean(tertiary revisions$age of patient),
             HFRS Band = unique(tertiary revisions$HFRS Band),
             IMD score = mean(tertiary revisions$IMD score),
             FinY = unique(tertiary_revisions$FinY),
             CV12mo = mean(tertiary revisions$CV12mo),
             TV12mo = mean(tertiary revisions$TV12mo),
             infection = unique(tertiary_revisions$infection))
# Align the levels of ProvCode in new data to match the training data
new data$ProvCode <- factor(new data$ProvCode, levels =
levels(tertiary_revisions$ProvCode))
# Align the levels of all relevant categorical variables
```

60

2	
3	new_data\$HERS_Band <- factor(new_data\$HERS_Band_levels =
4	lovels(tertiary, rovisions\$HEPS, Band))
5	
6	new_data\$sex <- factor(new_data\$sex, levels = levels(tertiary_revisions\$sex))
7	new_data\$FinY <- factor(new_data\$FinY, levels = levels(tertiary_revisions\$FinY))
8	new_data\$infection <- factor(new_data\$infection, levels =
9	levels(tertiary revisionsSinfection))
10	
11	#Easters are consistent with model
12	
13	
14	levels(new_data\$HFRS_Band)
15	levels(tertiary_revisions\$HFRS_Band)
17	
17	levels(new data\$ser)
19	lovels(hew_data;sex)
20	levels(tertiary_revisionspsex)
21	
22	levels(new_data\$FinY)
23	levels(tertiary revisions\$FinY)
24	
25	levels(new_data\$ProvCode)
26	levels(hew_data\$110veode)
27	levels(tertiary_revisions\$ProvCode)
28	
29	levels(new_data\$infection)
30	levels(tertiary revisions\$infection)
31	\sim
32	# Check levels of ProvCode in both datasets
33	satdiff(lovels(now, data\$ProvCodo) lovels(tertiary, rovisions\$ProvCodo)) #Lovels in
34	setulii(levels(ilew_ualasProvCoue), levels(tertiary_revisionssProvCoue)) # Levels in
35	new_data but not in tertiary_revisions
20 27	setdiff(levels(tertiary_revisions\$ProvCode), levels(new_data\$ProvCode)) # Levels in
38	tertiary_revisions but not in new_data
30	
40	new data\$ProvCode <- droplevels(new data\$ProvCode)
41	# Check for missing values in factor variables
42	
43	sum(is.na(new_data\$ProvCode)) # Number of missing values in ProvCode
44	
45	# Ensure that ProvCode is a factor
46	new data\$ProvCode <- factor(new data\$ProvCode, levels =
47	levels(tertiary_revisions\$ProvCode))
48	
49	
50	# Now try the prediction again
51	predicted_probs <- predict(model_with_custom_splines, newdata = new_data, type =
52	"response")
53	
54	
55	
56	
5/	
58	# Combine mean_unit_range and predicted_probs into a data frame
צנ	

2	
3	
4	
5	
6	
7	
/	
8	
9	
10	
11	
12	
12	
13	
14	
15	
16	
17	
18	
10	
19	
20	
21	
22	
23	
24	
27	
25	
26	
27	
28	
29	
30	
21	
21	
32	
33	
34	
35	
36	
37	
20	
38	
39	
40	
41	
42	
43	
11	
44 15	
45	
46	
47	
48	
49	
50	
50 E 1	
21	
52	
53	
54	
55	
56	
57	
57	
58	
59	

60

1

plot_data <- data.frame(OffPeakDriveDistanceMiles = DistanceMiles_range, predicted_prob = predicted probs) #Calculate 95% confidence intervals # Obtain predicted values and standard errors for the new data predictions <- predict(model_with_custom_splines, newdata = new_data, type = "link", se.fit = TRUE) # Calculate the confidence intervals for the log-odds scale (link scale) # Use a 95% confidence level (z-value = 1.96 for a 95% CI) z value <- 1.96 log odds lower <- predictions\$fit - z value * predictions\$se.fit log odds upper <- predictions\$fit + z value * predictions\$se.fit # Convert the log-odds confidence intervals to probabilities # First, apply the inverse link function (logistic function) to the log-odds lower_prob <- plogis(log_odds_lower)</pre> upper_prob <- plogis(log_odds_upper)</pre> # Combine the predicted probabilities and their confidence intervals into a data frame plot_data <- data.frame(</pre> DistanceMiles = new data\$OffPeakDriveDistanceMiles, predicted_prob = plogis(predictions\$fit), # Logistic transformation of the link ci lower = lower prob, ci upper = upper prob) library(ggplot2) # Plot the spline curve with confidence intervals ggplot(plot data, aes(x = DistanceMiles)) + geom line(aes(y = predicted prob), color = "blue", size = 1) + geom_ribbon(aes(ymin = ci_lower, ymax = ci_upper), fill = "blue", alpha = 0.2) + labs(x = "Distance (Miles)", y = "Predicted Probability Readmission at 30 days") + theme minimal() library(dplyr) # Group by mean_unit and calculate mean predicted_prob and corresponding confidence intervals mean data <- plot data %>% group by(DistanceMiles) %>% summarise(mean predicted prob = mean(predicted prob, na.rm = TRUE),

```
mean ci lower = mean(ci lower, na.rm = TRUE),
               mean ci upper = mean(ci upper, na.rm = TRUE)
              )
             # Define specific breaks (e.g., 25, 50, 75, ..., up to the maximum)
10
             breaks_seq <- seq(0, max(mean_data$DistanceMiles, na.rm = TRUE), by = 5)</pre>
11
12
13
             library(ggplot2)
14
             # Plot with specified increments on x-axis
15
             ggplot(mean data, aes(x = DistanceMiles, y = mean predicted prob)) +
16
              geom point() + # Add points for mean predicted prob
17
18
              geom line() + # Connect points with a line
19
              geom ribbon(aes(ymin = mean ci lower, ymax = mean ci upper), fill = "blue", alpha =
20
             0.2) + # Add ribbon for confidence intervals
21
              labs(x = "Off Peak Drive Distance Miles", y = "Mean Predicted Probability for Prolonged
22
             LOS", title = "Spline curve predicted probability of prolonged LOS by patient driving
23
24
             distance") +
25
              scale x continuous(limits = c(0, max(mean data$DistanceMiles, na.rm = TRUE)), breaks =
26
             breaks seq) +
27
              theme minimal() +
28
29
              theme(
30
               axis.title.x = element text(size = 14), # Increase x-axis title font size
31
               axis.title.y = element_text(size = 14), # Increase y-axis title font size
32
               axis.text.x = element_text(size = 12), # Increase x-axis tick label font size
33
               axis.text.y = element text(size = 12), # Increase y-axis tick label font size
34
35
               plot.title = element_text(size = 16, hjust = 0.5) # Increase plot title font size and center it
36
              )
37
38
39
40
41
42
             #Exposure 3 - PeakDriveTime
43
44
45
46
47
             library("Ime4")
48
49
             # Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for
50
             clustering
51
52
             m3.mi <- with(tertiary revisions, glm(Long Los ~ PeakDriveTime + IMD score + HFRS Band
53
             +
54
                                    sex + age of patient + infection + TV12mo + CV12mo + FinY +
55
             ProvCode,
56
57
                                   family = "binomial"))
58
59
60
```

3	
5	
4	
5	
6	
7	
ç	
0	
9	
10	
11	
12	
12	
15	
14	
15	
16	
17	
10	
10	
19	
20	
21	
22	
22	
23	
24	
25	
26	
27	
20	
20	
29	
30	
31	
32	
22	
27	
34	
35	
36	
37	
38	
20	
39	
40	
41	
42	
43	
11	
44	
45	
46	
47	
48	
10	
77	
50	
51	
52	
53	
51	
54	
55	
56	
57	
58	
50	
55	
60	

1 2

print(m3.mi)

Pool results across imputed datasets pooled results <- pool(m3.mi)</pre> # Summarize pooled results with confidence intervals summary pooled <- summary(pooled results, conf.int = TRUE) # Add Odds Ratios to the summary summary pooled\$OR <- exp(summary pooled\$estimate) summary pooled\$Lower Cl <- exp(summary pooled\$`2.5 %`) summary_pooled\$Upper_Cl <- exp(summary_pooled\$`97.5 %`)</pre> # Display the final table with Odds Ratios and Confidence Intervals print(summary_pooled) #check for evidence of multicollinearity? library(car) # Use the first imputed dataset for the VIF calculation complete data <- complete(tertiary revisions, 1) # Fit a logistic regression model on the complete dataset vif model <- glm(Read30days ~ DistanceMiles + IMD score + HFRS Band + sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode, data = complete data, family = "binomial") # Calculate VIF vif_values <- vif(vif_model)</pre> print(vif values) #Is there evidence of non linearity? # Custom function to add log-transformed variable and interaction term add interaction <- function(data) { data\$Log PeakDriveTime <- log(data\$PeakDriveTime) # Add log-transformed variable data\$Interaction <- data\$PeakDriveTime * data\$Log_PeakDriveTime # Add interaction term return(data) } # Extract the long-format data including the original data tertiary_revisions_modified <- complete(tertiary_revisions, action = "long", include = TRUE)</pre>

<pre># Apply the transformation to each imputed dataset tertiary_revisions_modified <- do.call("rbind",</pre>
Convert back to mids object tertiary_revisions_modified <- as.mids(tertiary_revisions_modified)
Fit the logistic regression model with the interaction term model <- with(tertiary_revisions_modified, glm(Long_Los ~ PeakDriveTime + Interaction, family = binomial(link = "logit")))
Pool the results pooled_results <- pool(model)
Summarize pooled results summary_pooled <- summary(pooled_results, conf.int = TRUE)
Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]
Print the p-value print(box_tidwell_p)
#P value 0.000916
#AIC of non spline model
model <- glm(Long_Los ~ PeakDriveTime, data = tertiary_revisions, family = binomial) summary(model)
#AIC 52843
<pre># Define a function to fit and evaluate spline models with knots based on centiles evaluate_centile_splines <- function(centiles, data) { # Calculate knots based on the specified centiles knots <- quantile(data\$PeakDriveTime, probs = centiles, na.rm = TRUE)</pre>
Fit a logistic regression model with natural splines using the calculated knots model_spline <- glm(Long_Los ~ ns(PeakDriveTime, knots = knots), family = binomial(link = "logit"), data = data)
Summarize the model summary_model <- summary(model_spline)

Extract p-values for the spline terms p_values <- summary_model\$coefficients[-1, "Pr(> z)"] # Exclude the intercept
Print the results cat("\nResults for centiles", centiles, ":\n") print(p_values)
<pre># Return the model and calculated knots for further inspection if needed return(list(model = model_spline, p_values = p_values, knots = knots)) }</pre>
Example centile configurations for 3, 4, and 5 knots centiles_3_knots <- c(0.05, 0.50, 0.95) # 5th, 50th, and 95th percentiles centiles_4_knots <- c(0.05, 0.35, 0.65, 0.95) # Custom centiles for 4 knots centiles_5_knots <- c(0.05, 0.25, 0.50, 0.75, 0.95) # 5 knots centiles
<pre># Evaluate models with centile-based knots using your dataset results_3_knots <- evaluate_centile_splines(centiles = centiles_3_knots, data = tertiary_revisions) results_4_knots <- evaluate_centile_splines(centiles = centiles_4_knots, data = tertiary_revisions) results_5_knots <- evaluate_centile_splines(centiles = centiles_5_knots, data = tertiary_revisions)</pre>
<pre># Compare models with centile-based knots cat("\nComparing models with different centile-based knots:\n") anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq")</pre>
<pre># Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots)</pre>
#52715, model with four knots best fit and significant spline terms and most parsimonious
<pre>#Run spline model with adjusted data excluding missing data library(splines) # For example, let's say you want 3 knots at specific percentiles knots <- quantile(tertiary_revisions\$PeakDriveTime, probs = c(0.05, 0.35, 0.65, 0.95), na.rm = TRUE) print(knots)</pre>
<pre>spline_terms <- ns(tertiary_revisions\$PeakDriveTime, knots = knots)</pre>

model_with_custom_splines <- glm(Long_Los ~ ns(PeakDriveTime, knots = HFRS_Band + IMD_score + sex + age_of_patient + infection + TV12mo + CV12mo + family = "binomial", data = tertiary_revisions)	knots) + FinY + ProvCode,
summary(model_with_custom_splines)	
#Generate a sequence of mean unit values for predicting	
DistanceMiles_range <- seq(min(tertiary_revisions\$PeakDriveTime), max(tertiary_revisions\$PeakDriveTime), length.out = 100)	
<pre>new_data <- expand.grid(PeakDriveTime = DistanceMiles_range, sex = levels(tertiary_revisions\$sex), # Ensure it takes all factor levels age_of_patient = mean(tertiary_revisions\$age_of_patient, na.rm = TRUE), HFRS_Band = levels(tertiary_revisions\$HFRS_Band), # Ensuring correct fact IMD_score = mean(tertiary_revisions\$IMD_score, na.rm = TRUE), FinY = levels(tertiary_revisions\$FinY), # Ensuring correct factor levels CV12mo = mean(tertiary_revisions\$CV12mo, na.rm = TRUE), TV12mo = mean(tertiary_revisions\$TV12mo, na.rm = TRUE), ProvCode = levels(tertiary_revisions\$ProvCode), # Ensuring correct factor infection = levels(tertiary_revisions\$infection) # Ensuring correct factor levels)</pre>	ctor levels levels evels
# Align the levels of ProvCode in new_data to match the training data new_data\$ProvCode <- factor(new_data\$ProvCode, levels = levels(tertiary_revisions\$ProvCode))	
<pre># Align the levels of all relevant categorical variables new_data\$HFRS_Band <- factor(new_data\$HFRS_Band, levels = levels(tertiary_revisions\$HFRS_Band)) new_data\$sex <- factor(new_data\$sex, levels = levels(tertiary_revisions\$se new_data\$FinY <- factor(new_data\$FinY, levels = levels(tertiary_revisions\$ new_data\$infection <- factor(new_data\$infection, levels = levels(tertiary_revisions\$infection))</pre>	ex)) FinY))
#Factors are consistent with model	
levels(new_data\$HFRS_Band) levels(tertiary_revisions\$HFRS_Band)	
levels(new_data\$sex)	

2	
3	lovels (tortion, revisions (cov)
4	levels(tertiary_revisions\$sex)
5	
6	levels(new_data\$FinY)
7	levels(tertiary revisions\$FinY)
8	
9	lough (nour detection of the code)
10	
11	levels(tertiary_revisions\$ProvCode)
12	
13	levels(new data\$infection)
14	levels(tertiary_revisions\$infection)
15	
16	If Charles the of Data Carda in the determined
17	# Check levels of ProvCode in both datasets
18	setdiff(levels(new_data\$ProvCode), levels(tertiary_revisions\$ProvCode)) # Levels in
19	new_data but not in tertiary_revisions
20	setdiff(levels(tertiary_revisions\$ProvCode), levels(new_data\$ProvCode)) # Levels in
21	tertiary revisions but not in new data
22	
23	
24	new_data\$ProvCode <- droplevels(new_data\$ProvCode)
25	# Check for missing values in factor variables
20	sum(is.na(new_data\$ProvCode)) # Number of missing values in ProvCode
27	
20	# Ensure that ProvCode is a factor
30	
31	new_data\$ProvCode <- factor(new_data\$ProvCode, levels =
32	levels(tertiary_revisions\$ProvCode))
33	
34	# Now try the prediction again
35	predicted probs <- predict(model with custom splines newdata = new data type =
36	"response")
37	response)
38	
39	
40	
41	
42	# Combine mean unit range and predicted probs into a data frame
43	# combine mean_unit_range and predicted_probs into a data mane
44	piot_data <- data.frame(PeakDriveTime = DistanceWilles_range, predicted_prob =
45	predicted_probs)
46	
47	#Calculate 95% confidence intervals
48	
49	# Obtain prodicted values and standard errors for the new data
50	
51	predictions <- predict(model_with_custom_splines, newdata = new_data, type = "link",
52	se.tit = TRUE)
53	
54 55	# Calculate the confidence intervals for the log-odds scale (link scale)
55 56	# Use a 95% confidence level (z-value = 1.96 for a 95% CI)
57	\sim volue \sim 1.06
58	
59	log_oads_lower <- predictions\$fit - z_value * predictions\$se.fit
60	log_odds_upper <- predictions\$fit + z_value * predictions\$se.fit
~~	

```
# Convert the log-odds confidence intervals to probabilities
# First, apply the inverse link function (logistic function) to the log-odds
lower prob <- plogis(log odds lower)</pre>
upper_prob <- plogis(log_odds_upper)</pre>
# Combine the predicted probabilities and their confidence intervals into a data frame
plot data <- data.frame(
 DriveTime = new data$PeakDriveTime,
 predicted_prob = plogis(predictions$fit), # Logistic transformation of the link
 ci lower = lower prob,
 ci upper = upper prob
)
library(ggplot2)
# Plot the spline curve with confidence intervals
ggplot(plot data, aes(x = DriveTime)) +
 geom line(aes(y = predicted prob), color = "blue", size = 1) +
 geom_ribbon(aes(ymin = ci_lower, ymax = ci_upper), fill = "blue", alpha = 0.2) +
 labs(x = "Distance (Miles)", y = "Predicted Probability Readmission at 30 days") +
 theme_minimal()
library(dplyr)
# Group by mean unit and calculate mean predicted prob and corresponding confidence
intervals
mean data <- plot data %>%
 group_by(DriveTime) %>%
 summarise(
  mean predicted prob = mean(predicted prob, na.rm = TRUE),
  mean ci lower = mean(ci lower, na.rm = TRUE),
  mean ci upper = mean(ci upper, na.rm = TRUE)
 )
# Define specific breaks (e.g., 25, 50, 75, ..., up to the maximum)
breaks seq <- seq(0, max(mean data$DriveTime, na.rm = TRUE), by = 5)
library(ggplot2)
# Plot with specified increments on x-axis
ggplot(mean_data, aes(x = DriveTime, y = mean_predicted_prob)) +
 geom point() + # Add points for mean predicted prob
 geom line() + # Connect points with a line
```

```
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
```

```
geom_ribbon(aes(ymin = mean_ci_lower, ymax = mean_ci_upper), fill = "blue", alpha =
0.2) + # Add ribbon for confidence intervals
 labs(x = "Peak Drive Times (Minutes)", y = "Mean Predicted Probability for Prolonged LOS",
title = "Spline curve predicted probability of prolonged LOS by patient driving times") +
 scale x continuous(limits = c(0, max(mean data$DriveTime, na.rm = TRUE)), breaks =
breaks seq) +
 theme minimal() +
 theme(
            = 1.
.e = 14),
.ize = 12), *
.(size = 12), # h
.(size = 16, hjust = 0
  axis.title.x = element_text(size = 14), # Increase x-axis title font size
  axis.title.y = element text(size = 14), # Increase y-axis title font size
  axis.text.x = element text(size = 12), # Increase x-axis tick label font size
  axis.text.y = element text(size = 12), # Increase y-axis tick label font size
  plot.title = element_text(size = 16, hjust = 0.5) # Increase plot title font size and center it
 )
```

####END####

infection

Spell_Los

OffPeakDriveDistanceMiles

008808.

TV12mo

PeakDriveTime

BMJ Open

What is the impact of longer patient travel distances and times on perioperative outcomes following revision knee replacement: A retrospective observational study using data for England from Hospital Episode Statistics

Journal:	BMJ Open
Manuscript ID	bmjopen-2024-085201.R3
Article Type:	Original research
Date Submitted by the Author:	10-Mar-2025
Complete List of Authors:	Matthews, Alexander; University of Exeter Faculty of Health and Life Sciences, Department of Clinical and Biomedical Sciences; Royal Devon and Exeter NHS Foundation Trust, Evans, Jonathan; University of Exeter Medical School, Health Services and Policy Research; Royal Devon and Exeter NHS Foundation Trust, Princess Elizabeth Orthopaedics Centre Evans, Jonathan; University of Exeter, Department of Public Health and Sports Sciences Lamb, Sarah; University of Exeter, College of Medicine and Health Price, Andrew; University of Oxford, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences Gray, William; NHS England and NHS Improvement London Briggs, Tim; NHS England and NHS Improvement London Toms, Andrew ; Princess Elizabeth Orthopaedic Centre
Primary Subject Heading :	Public health
Secondary Subject Heading:	Surgery
Keywords:	Knee < ORTHOPAEDIC & TRAUMA SURGERY, PUBLIC HEALTH, Health Services

SCHOLARONE[™] Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our <u>licence</u>.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which <u>Creative Commons</u> licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

terez oni

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies



Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

Title What is the impact of longer patient travel distances and times on perioperative outcomes following revision knee replacement: A retrospective observational study using data for England from Hospital **Episode Statistics** Names, Affiliations, and positions of all authors Alex Matthews1,2,3,4, Jonathan P Evans2,3, Jonathan T Evans2,3, Sarah E Lamb3 Andrew Price4,5, William Gray1, Tim Briggs1,6, Andrew Toms2,3 1. Getting It Right First Time programme, NHS England, London, UK 2. Royal Devon University Healthcare NHS Foundation Trust, Exeter, UK 3. University of Exeter, Exeter, UK 4. Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, UK

2		
3 4	21	5. Nuffield Orthopaedic Centre, Oxford, UK
5 6	22	6. Royal National Orthopaedic Hospital, Stanmore, London, UK
7 8	23	Correspondence: Alexander Matthews, Getting It Right First Time programme,
9 10	24	NHS England and NHS Improvement, Wellington House, Waterloo Road. London,
11 12	25	UK, SE1 8UG; telephone: 07401932627 email: alexander.matthews1@nhs.net
13 14 15	26	Word count: [3316 words]
16 17	27	Short title: Patient Journey Distance and Time for Complex Revision Knee Surgery
18 19 20	28 29	Key words: social determinants of health, revision knee replacement, contextual factors in complex health interventions
21 22	30 31	Twitter handles:
23		
24 25 26	32	Alex Matthews: @DrAlexMatthews
20 27 28 29 30	33	Tim WR Briggs: @ProfTimBriggs
	34	WilliamKGray1: @WilliamKGray1
31 32 22	35	GIRFT programme: @NHSGIRFT
33 34 35	36	
36	37	
37 38	38 39	Structured Abstract
39 40 41	40	Objectives
42	41	Patients undergoing revision total knee replacement (RevKR) surgery often have
43 44	43	difficulties mobilising and increasingly rely on family support. Evolving practice in
45 46	44	England aims to manage these patients in specialised centres with the intention of
46 47 48 49	45	improving outcomes. This practice will result in longer travel distances and times in
	46	this frailer group of patients. We want to examine the types of distances and travel
50 51	47	times patients can be expected to travel for this complex orthopaedic surgery and to
52 53	48	explore concerns of how these impact patient outcomes.
54		
55 56		
57 58		
59		

2		
3 ⊿	49	Design
5	50	
6 7	51	Retrospective observational study from the Hospital Episode Statistics. Multivariable
8	52	adjusted logistic regression models were used to investigate the relationship
9 10	53	between patient travel distances and times with perioperative outcomes.
11	54	Setting
12 13	54 55	Setting
14	56	Patients presenting to tertiary referral centres between 1 st January 2016 to 31 st
15 16	57	December 2019. A tertiary referral centre was defined as a trust performing >49
17 18	58	revisions in the year prior.
19 20	59	Participants
21	60	
22 23	61	Adult patients undergoing RevKR procedures for any reason between 1 st January
24	62	2016 to 31 st December 2019.
25 26	63	Exposure
27	64	
28 29	65	The shortest patient level travel distance and time was calculated using the
30 21	66	department of health Journey Time Statistics using TRACC software and Dijkstra's
32 33	67	algorithm.
34 35	68	Main Outcome Measures
36 27	69 70	The primary outcome is emergency readmission within 30 days. Secondary
37 38	71	outcomes are mortality within 90 days and length of inpatient stay
39 40	/1	
40 41	72	Results
42 43	73	6.000 potients underwant Day/KD at 26 tertiany referral control. There was a weak
43 44	/4	6,000 patients underwent RevRR at 56 tertiary referral centres. There was a weak
45 46	75	correlation between social deprivation and travel distance, with patients from the
47	76	most deprived areas travelling longer distances. Overall, 30-day readmission was
48 49	77	not statistically associated longer driving distance (OR 1.00 95% CI 0.99 to 1.02) or
50 51	78	peak driving times (OR 1.00 95% CI 0.99 to 1.01).
52	79	Conclusions
55 54	80	
55	81	There was no association between increasing travel distance and time on
56 57 58	82	perioperative outcomes for RevKR patients.
58 59	83	
60	84	

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

85	
86	
87	
88	
00 00	
09	
90	
91	
92	
93	
94	
95	
06	
90	
97	
98	
99	
100	
101	Strengths and limitations of this study
102	• This study is one of the largest studies in the literature investigating outcomes
102	
103	following revision knee replacement.
104	This data reflects revision know replacement procedures undertaken seress
104	• This data reliects revision knee replacement procedures undertaken across
105	different geographical areas of England
100	
107	Outing to differences in the course of Heavital Eniords Otatistics
106	• Owing to differences in the coverage of Hospital Episode Statistics,
107	procedures in hospitals outside of England were not included in this analysis
107	
100	Oliviaal adding any sting is language to some truste with some truste many
108	Clinical cooling practice is known to vary across trusts, with some trusts more
109	consistent in coding than others which may have created some bias in the
107	
110	model estimates
111	This analysis only reports travel times for natients with access to their own
111	• This analysis only reports traver times for patients with doces to their own
112	transport and does not consider times for those patients using public transport
113	
115	
114	
115	
116	
	 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116
Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

117 Introduction

Primary knee replacement is a successful procedure that improves quality of life for the majority of patients.[1] However, at 10 years following a primary knee replacement, about 3.5% of patients will have undergone a revision surgery.[2] The majority of these procedures are carried out due to infection or polyethylene wear of the implant.[3] A failed primary knee replacement represents a life changing transition point where individuals are likely to suffer from pain, reduced mobility as well as dependency on family members. [4] Patients often face multi-step surgery with longer hospital length of stays and higher complication rates [5, 6]

The Getting It Right First Time (GIRFT) programme orthopaedic National Report was published in 2015.[7] A key recommendation was the centralisation of complex orthopaedic surgery, including revision knee surgery, to specialist centres with the aim of improved patient outcomes. Consequently, revision total knee replacement (RevKR) surgery in the England has evolved into a regional network service model.[8] All hospitals performing RevKR form a network in the respective regions. Less specialist hospitals, defined by lower annual case volume thresholds, are encouraged to discuss and sometimes refer their caseload to more specialist centres. Several studies based on large revision hip and knee registries have suggested this model carries a lower failure rate defined by the need for further revision surgery.[9-11] Early evidence has suggested reduced early failure rates through the adoption of revision knee networks.[12]

43 140

However, for some patients, this approach to managing patients is inevitably associated with increasing travel distances between patient's homes and their treating hospital. Travel distance has been shown to be an important factor in patient choice when selecting a surgeon for joint replacement surgery. It may be even more important for those awaiting revision joint replacement surgery as these patients struggle with mobility, may be unable to drive and may be more reliant on family members.[4] Evidence suggests that patients considering joint replacement are prepared to travel longer distances to obtain the best possible outcomes. A requisite in making such a decision requires data on outcomes of patients travelling greater

distances. Patients travel longer distances have been found to have higher readmission rates and higher mortality rates when undergoing other types of specialised surgery.[13] The pick-up rate of early complications, avoiding the need for readmission, may be less in areas further away from the main treatment centre. There is also concern that patients required to travel greater distances are more likely to be re-admitted to a different hospital than that where surgery was undertaken, resulting in clinical decisions that do not incorporate the primary surgeon and so potentially leading to poorer outcomes. [14] There is an absence of evidence in the literature to support or refute this argument in the context of patients undergoing RevKR. Therefore the aim of this paper is to investigate the relationship between longer patient travel distances and perioperative outcomes following RevKR performed in high volume tertiary referral centres.

163 Methods

165 Design

1166This study is a retrospective data analysis of observational data from the Hospital2167Episode Statistics (HES) and Office for National Statistics (ONS) databases. HES4168data is collected by NHS England for all patients treated at NHS hospitals in England6169and those treated at private hospitals where treatment was funded by the NHS. This7170study complies with the recommended reporting guidelines when using HES data[15]9171and the Strengthening of Reporting of Observational studies in Epidemiology172(STROBE) guidelines.[16]

The analysis and presentation of data follows current NHS England guidance for the use of HES data for research purposes[17] and is anonymised to the level required by ISB1523 Anonymisation Standard for Publishing Health and Social Care Data.[18] The HES data were linked at a patient level to data from the ONS on deaths and date of death, which allowed the identification of patients who had died after their surgery. Linkage was achieved using a unique pseudonymised patient identifier using a previously validated methodology.[19] Patient travel distances were calculated using the Journey Time Statistics reference

⁵⁷ 180 Patient traver distances were calculated using the 300 mey time statistics reference
 ⁵⁸ 181 document produced by the UK Department of Transport which modelled theoretical
 ⁵⁹ 182 journey times between known centroids of Lower Layer Super Output Areas (LSOA)

183 of residence and NHS hospital sites.[20] Please refer to Supplementary material
184 S1 for Journey Times Statistics reference document.

186 Population

An RevKR procedure was defined as a permanent removal or exchange of knee arthroplasty components. This includes a revision of a total knee replacement and a conversion of a unicondylar knee replacement to a total knee replacement. Secondary patellar resurfacing was not included as this represents a simple revision procedure, one that can be carried out in most nonspecialised hospitals. All patients aged \geq 18 years who underwent a RevKR in a high-volume trust between 1st January 2016 to 31st December 2019 were included in the study population. A high-volume trust was classified as a centre performing > 49 revisions per year. This revision volume threshold for classification represents that proposed by the British Association for Surgeons of the Knee (BASK) Revision Knee Working Group and is a mandatory requirement for all highly specialist centres co-ordinating regional networks. [21] As such centres attaining this threshold are more likely to represent tertiary referral centres where the stratification of more complex work will take place. Annual case volume at each trust was defined as the number of revision cases conducted in the year prior to the index procedure. This measure was preferred over a simple calculation of average annual volume as it accounts for recent experience at the point of surgery. The Office for Population Censuses and Surveys' Classification of Interventions and Procedures version 4 (OPCS-4) codes used to identify RevKR procedures are detailed in **Supplementary material S2**. Since laterality was needed to identify re-revisions, patients were excluded where the procedure laterality was not specified. The flow of patients, with numbers excluded at each point, is summarised in Supplementary material S3. To manage population heterogeneity, data were extracted for the period 1st April 2011 to 31st December 2019 and only the first revision for a specific side of the body record in this time period included. [22] Thus, any early revisions on the same side of the body in the four years and nine months preceding the start of the study period were identified and these patients excluded from the study. This aims to exclude the early revision knee replacement failures which have been shown to represent catastrophic failures potentially skewing our results.[22] We included revisions for infection as, despite

these representing a more variable patient group, presence of infection was thought to be unrelated to how far a patient lives from a specialised referral centre.

Exposure variable

Travel distances and times were calculated between a patient's LSOA and the postal codes for the treating hospitals. LSOAs are determined by the Office for National Statistics and are designed for the reporting of small area statistics. Public transport and highways data for England were used to create theoretical journey distances and times from origins to destinations. A network of journey distances and times from origins to destinations was produced using a software package called Transport Accessibility and Connectivity Calculator (TRACC). The Dijkstra's algorithm calculated the shortest route between these points. Data linkage between the HES/ONS dataset and the travel times dataset was achieved using two shared data fields; LSOA and hospital site. The resulting travel distances and/or times for each patient were analysed as continuous variables. Three exposure variables were used. Straight line travel distance represented the distance "as the crow flies" between a patient's LSOA and treating hospital. Off peak driving distance represented the shortest driving distance between a patients LSOA and treating hospital. Finally peak driving times were calculated using average traffic speeds between 7am and 10am for the shortest possible road route between a patients LSOA and treating hospital. These three variables were used to account for variation in travel infrastructure between rural and urban areas and to attribute more meaningful results for patients.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Co-variates and cluster variable

The following groups of known or potential confounding variables were chosen a priori for inclusion in our multivariable logistic regression modelling:

Patient factors: Age in years (continuous), sex (male/female). Health co-morbidity was quantified using the Hospital Frailty Risk Score (HFRS). HFRS identifies frailty based on the occurrence of any of 109 International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10) codes used during any hospital admissions in the two years prior to, and for, the index admission. Deprivation was measured using the Index of Multiple Deprivation (IMD).[23] The

1

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

2 3	250	IND since the LOOA where the national lines a second here a near of management
4	250	IMD gives the LSOA where the patient lives a score based on a range of measures
5 6 7 8	251	of deprivation. IMD was analysed as a continuous variable.
	252	Clinical factors: Defined by the presence or absence of infection as the primary
9	253	indication for RevKR. This was identified from the International Statistical
10 11	254	Classification of Diseases and Related Health Problems, tenth revision (ICD-10)
12 13	255	codes used during the admission.
14	256	Surgical factors: Surgeon and hospital volume (both continuous) was defined as the
16 17	257	number of RevKRs performed by a consultant or hospital in the 365 days prior to
18 19 20 21	258	each index procedure across the entire cohort. This was calculated before any
	259	exclusion criteria was applied.
22 23	260	Temporal factors: Financial year of procedure (2015/16, 2016/17, 2017/18, 2018/19,
24 25 26	261	2019/20).
26 27	262	Hospital Provider: Clustering of patients by hospital provider was initially modelled
28 29 30 31 32 33 34 35 36 37 38 39 40	263	using random effects. However, despite variability between hospital providers with
	264	primary and secondary outcomes, instability in the model estimates were observed.
	265	To address the possibility of clustering at this level, a fixed effects model was
	266	adopted with hospital provider as a covariate.
	267	
	268	
	269	Outcomes
	270 271	The primary outcome was emergency readmission within 30 days of discharge from
42	272	the index surgical hospital. Readmission in this early period is very likely related to a
43 44	273	complication of the surgical procedure. It has been used as a marker of perioperative
45 46	274	outcomes in similar studies investigating the relationship between patient travel
47 48	275	distance and outcomes following surgery. [13]
49 50 51	276	Secondary outcomes were:
52	277	90-day all-cause mortality, identified using linked data from Civil Registrations
53 54	278	(Mortality) dataset;
55 56 57	279	Inpatient length of hospital stay was attributed from continuous inpatient spells
58	280	(CIPS), which is the preferred estimate of length of stay. This refers to the length of
59 60	281	first stay after the operation regardless of any transfers across providers. The

BMJ Open

2 3	202	
4 5 6 7 8 9 10	282	median length of stay was calculated after visually inspecting the distribution and this
	283	was dichotomized into prolonged length of stay it longer than the median stay.
	284	
	285 286	Statistical Analyses
11	280 287	Data was extracted from a secure, encrypted server controlled by NHS England.
12 13	288	Data were analysed within a secure, encrypted environment using standard
14 15	289	statistical software: R Studio version 2023.09.1+494 (Boston, Massachusetts, USA).
16 17	290	The R code and packages used are included in Supplementary material S4 .
18 19	291	Missing data were managed according to its extent and relevance to the aims of this
20 21 22	292	study. Age and IMD score were imputed for the small number of missing cases using
	293	the mean of the entire study cohort. Given the central role of LSOA in estimating
23 24	294	travel distances and times and fewer than 5% of cases with missing data, these
25 26	295	cases were excluded to avoid the introduction of bias. Following data linkage
27 28 29 30 31 32 33 34 35 36 37 38	296	between the HES/ONS dataset and the travel times dataset, approximately 36% (n =
	297	5,838) of cases did not match. Multiple imputation was performed using predictive
	298	mean matching based on the entire cohort of patients with the following predictors:
	299	age, sex, HFRS score, IMD score, hospital provider code, hospital volume and
	300	surgeon volume. Dependent variables including readmission at 30 days, mortality at
	301	90 days and length of stay were also used in the imputation following a
	302	recommended approach using preditive mean matching[24]. A total of five
39 40	303	imputations were randomly chosen and subsequent regression analyses were
41 42	304	performed.[25] Imputed data is shown in Supplementary material S5 .
43 44	305	Patient travel distances were categorised into quintiles for interpretation of baseline
45 46	306	demographics and clinical characteristics. Subsequent analysis of travel distances
47	307	and times were performed as continuous variables. Spearman's rank correlation was
48 49	308	performed to investigate the relationship between IMD score and patient age with
50 51	309	travel distances.
52 53	310	Straight line travel distance was modelled with restricted cubic splines to allow for
54 55	311	the non-linear effects when testing the association with the primary outcome. All
56 57	312	exposures were modelled with restricted cubic splines to allow for the non-linear
58	313	effects when testing the association with prolonged length of stay. The Akaike
59 60	314	Information Criterion was used to select the most parsimonious specification of

restricted cubic splines using the final adjusted model. Fixed effects logistic regression models were used for the outcomes of readmission at 30 days, mortality at 90 days and prolonged length of stay. Where implemented, the use of splines was used to create figures depicting the association between travel distance or times and probability of the outcomes. Only adjusted spline models were used to depict these associations. All co-variates were included in the adjusted models. Multicollinearity was assessed using eigenvalues, variance inflation factors and by examination of model parameter estimates with the unadjusted model. Odds ratios with 95% CIs and associated p-values were reported. A p-value of < 0.05 was taken to indicate statistical significance.

Results

Overview of results

A total of 16,736 patients met the inclusion criteria. Excluding missing LSOA data (n=171), 16,565 patients were included in the analysis. Following data linkage with department of transport journey times statistics, 10,727 patients had complete data linkage and data were imputed for the remaining 5,838 (35.2%). Of the 16,565 patients, 41.5% (n=6,880) presented to a tertiary referral centre and these data formed our analysis cohort. Patients were operated on across 181 hospital sites and 38 hospital trust providers. The baseline demographic and clinical characteristics of the patients were broadly similar between guintiles of straight-line travel distance. (Table 1). Higher hospital volumes were seen in patients travelling longer distances. Figure 1 shows that straight line travel distance was weakly correlated with age (r= -0.05, p value <0.05) and social deprivation(r = -0.05, p value <0.05). Older patients were less likely to travel farther distances. Patients from the least deprived areas travelled shorter distances.

> Table 1 – Baseline patient demographics and clinical characteristics stratified by travel distance quintiles from first imputed dataset

	Travel Distan	ce Quintile			
	1	2	3	4	5
Distance	2.09 (1.35	4.42 (3.91	7.08 (6.34 to	11.39 (10.11	22.42 (18.0
(Miles)	to 2.75)	to 5.00)	7.99)	to 12.74)	to 32.19)
Driving Time	13 (9.3 to	20.45 (17 to	26.30 (21.98	34.10 (29.68	52.05 (42.6
(Minutes)	17)	25)	to 31.13)	to 40.20)	to 66.83)
Number of	1376	1376	1376	1376	1376
patients					
Tertiary	37 (97.37%)	38 (100%)	36 (94.74%)	35 (92.11%)	37 (97.37%
Providers					
Age Mean	69.71	69.96	69.66 (10.92)	68.84 (11.01)	68.58 (10.7
(SD)	(10.81)	(10.71)			
Female Sex	762	768	729 (52.98%)	722 (52.47%)	734 (53.34
	(55.38%)	(55.81%)			
HFRS None	647	620	614 (44.62%)	666 (48.40%)	676 (49.13
	(47.02%)	(45.06%)			
HFRS Mild	438	474	485 (35.25%)	465 (33.79%)	433 (31.47
	(31.83%)	(34.45%)			
HFRS	241	236	243 (17.66%)	198 (14.39%)	230 (16.72
Moderate	(17.51%)	(17.15%)			
HFRS Severe	50 (3.63%)	46 (3.34%)	34 (2.47%)	47 (3.42%)	37 (2.69%)
Infection	314	331	310 (22.53%)	334 (24.27%)	355 (25.80
Present	(22.82%)	(24.06%)			
Surgeon	7 (3 to 13)	7 (3 to 13)	8 (3 to 15)	8 (3 to 16)	9 (4 to 17)
Volume					

Hospital	73 (60 to	74 (60 to	79 (63 to 97)	79 (63 to 99)	85 (68.75 to
Volume	87)	89)			112)
IMD Score	16.44 (8.73	14.30 (7.96	14.50 (8.47	14.83 (9.23	14.752 (8.78
	to 28.67)	to 24.57)	to 21.36)	to 21.74)	to 21.45)
Year 2015/16	104 (7.56%)	94 (6.83%)	94 (6.83%)	89 (6.47%)	92 (6.69%)
Year 2016/17	383	354	348 (25.29%)	338 (24.56%)	353 (25.65%)
	(27.83%)	(25.73%)			
Year 2017/18	384	365	339 (24.64%)	360 (26.16%)	336 (24.42%)
	(27.91%)	(26.53%)			
Year 2018/19	269	325	347 (25.22%)	354 (25.73%)	339 (24.64%)
	(19.55%)	(23.62%)			
Year 2019/20	236	238	248 (18.02%)	235 (17.08%)	256 (18.60%)
	(17.15%)	(17.30%)			
Outcomes					
The primary a	nd secondary	outcomes are	summarised in	n table 2.	
The observed	rate of readmi	ssion at 30 da	iys was 8.3% (568/6880). Th	ere was a
negative assoc	ciation betwee	n higher straig	ght line travel c	listances and e	emergency
readmission at	t 30 days (Figu	ure 2). Howev	er wide confide	ence intervals	precluded
statical inferen	ices. In additio	n, higher trave	el distance by	road and longe	er drive times
were not asso	ciated with sta	tistically worse	e readmission	rates at 30 day	/s. The rate of
mortality at 90	days was only	/ 3.2% (217/6	880). No statis	tically significa	nt relationship
was observed	between the c	listance a pati	ent travels by	road or the tim	e a patient
spends travelli	ng at peak dri	ving times with	n rates of mort	ality at 90 days	s. 49.7%
(3421/6880) of	f patients repo	rted hospital s	stays more that	n 5 days. Follo	wing

BMJ Open

2 3 4	367	adjustment of confounding factors, we observed no associations between prolonged					
5 6	368	length of stay and pat	tient travel distance (F	-igures 3-5)			
7 8	369						
9 10	370	Table 2 – Adjusted pooled Multivariable Logistic Regression showing Odds					
11 12 13 14 15 16 17 18 19	371 372	Ratios for primary a	nd secondary outco	mes by exposure va	ariables		
			Straight line travel	Travel distance by	Peak Travel times		
			distance (OR, 95%	shortest road route	by shortest road		
			CI)	(OR, 95% CI)	route (OR, 95% CI)		
20 21		Readmission with	Figure 2	1.00 (0.99 to 1.02), p	1.00 (0.99 to 1.01), p		
22 23		30 days		value = 0.81	value = 0.69		
24 25		90 Day Mortality	1.00 (0.98 to 1.02), p	1.00 (0.99 to 1.01), p	1.00 (0.99 to 1.01), p		
26 27 28			value = 0.87	value = 0.86	value = 0.89		
29		Prolonged Length of	Figure 3	Figure 4	Figure 5		
30 31 32		stay					
33 34 35 36 37 38 39 40 41 42	373 374 375 376	•Odds ratios have been	adjusted for patient age	, sex, HFRS score,			
	377 378 379	Discussion					
43 44 45	380	Statement of principal findings					
46	381	We present a multi-hospital site retrospective analysis of patients undergoing					
47 48	382	revision knee replace	ment surgery at tertia	ry referral centres in	England. In this		
49 50	383	analysis of 6,880 pati	ents undergoing Revl	KR, we did not observ	e a statistical		
51 52	384	association between distance and time travelled for revision surgery and					
53	385	readmission within 30) days.				
54 55 56 57 58 59 60	386 387	Strengths and weakn	esses of the study				

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

The findings of this study should be interpreted in view of several limitations. Firstly, this analysis used observational data from a large administrative dataset covering all NHS-funded procedures conducted in England. As with all administrative datasets we are limited in the amount of detail provided regarding presentation. We chose to categorise a high-volume centre by trust to accurately capture surgical experience. All NHS hospitals in England are run by hospital trusts which typically involve between one and four hospitals within a catchment area standardising their practice. It is common practice for specialist orthopaedic surgeons to move between these sites delivering the same procedures. Our study involved 187 hospital sites run by 38 trusts. We acknowledge this is a weakness of our study as this may not be representative of all trusts. We included all indications for RevKR in our patient cohort because indication was not thought to be related to how far a patient lives from a hospital. However, we acknowledge the rate of complications is higher in patients with infection and we subsequently adjusted for indication for revision in our analyses. [26] It is likely that because we did not exclude previous revision knee arthroplasty patients, the complexity of the surgery undertaken in our cohort varied. We recognise this is a limitation of the study however we assume case mix was unrelated patient travel distance.

There were many missing patients (approximately 36%) following the linkage of HES data with Journey Time Statistics. To account for this, assumed that the data was missing at random and used multiple imputation to estimate missing travel distances. It is likely the imputed values may introduce bias, however we modelled these based on predictors and dependent variables to improve our estimates. We do not present a sample size calculation, rather we have used all available data and our sample size was set by our inclusion criteria. We controlled for the clustered nature of our data between hospital providers through inclusion as a covariate in our modelling. To ensure consistency in our definition of tertiary referral hospitals, only hospitals performing >49 revisions/year were included. These are likely to treat a similar case mix of patients and potentially have similar access to resources within a national healthcare system. This approach allowed us to control for variation across providers. However, we acknowledge it does not fully account for the hierarchical

BMJ Open

1	
2	
3	
4	
5	
6	
0	
/	
8	
9	
10	
11	
12	
13	
14	
15	
16	
10	
1/	
18	
19	
20	
21	
22	
23	
24	
27	
25	
20	
27	
28	
29	
30	
31	
32	
33	
31	
25	
22	
36	
37	
38	
39	
40	
41	
42	
43	
44	
15	
45	
40	
4/	
48	
49	
50	
51	
52	
53	
54	
55	
22	

420 nature of the data with differences in treatment protocols and hospital specialisation 421 among factors which may influence patient outcomes.

423 There is a lack of granular clinical data using HES for each readmission. Therefore, 424 we cannot ascertain precise reasons for readmissions, but we assume are related to 425 a post-surgical complication. Information on the exact date of readmission and death 426 was also not available. Therefore, a time-to-event approach in outcome analysis was 427 not possible. Clinical coding practice within HES is known to vary across trusts.[27] 428 As an example, some trusts may be more consistent in coding comorbidities, and 429 this may have created some bias. However, this is unlikely to vary systematically 430 with travel distances and so significantly bias our findings. We acknowledge the 431 relatively short travel distances in this population compared to examples from the United States as such the results of this study may not be generalisable to larger 432 433 geographical areas or less mature healthcare systems. However, the upper quintile 434 in our study represents a substantial journey distance and time for our patient cohort 435 where poor mobility is a significant factor affecting their care. This analysis does not 436 consider journey times of those who may not have access to a car and instead 437 chose to take public transport.

438

441

Strengths and weaknesses in relation to other studies, discussing important 439 440 differences in results

This is the first study to analyse the potential impact of patient travel distances on 442 443 patients receiving RevKR. The findings that longer travel distances are not 444 associated with inferior outcomes is an important part of the evaluation of the 445 assumptions and context behind the establishment of revision knee networks.[28] 446 This study has shown that concerns of introducing a network in larger geographical 447 regions, for example in Scotland where longer patient travel distances and times are 448 common, may be less important. [29] This is particularly useful as regions explore the 449 geography of their revision networks and during summative outcome assessment of 56 450 this complex health intervention.[30] Despite there being a potential negative 57 58 451 association between straight line travel distance and emergency readmission at 30 59 60

Page 18 of 116

1 2		
3 4	452	days, there was a lack of association involving driving distances and times which
5	453	present real world challenges for patients.
7 8	454	It may be seen as surprising that no association between travel distance and
9 10	455	prolonged length of hospital stay was identified. An expectation exists of increasing
11	456	difficulties being encountered with the discharge of patients living greater distances
12 13	457	from their treating hospital, which has been observed in patients following elective
14 15 16 17 18	458	pancreatic surgery.[31] This is also an observation seen in patients being treated in
	459	specialist vascular centres in the United States which led to the recommendation of
	460	additional care coordination and follow up efforts. However, the geography of the
19 20	461	population in these studies was much larger with significantly longer travel distances.
21 22 23 24 25 26 27 28 29 30 31 32 22	462	We did observe a weak but statistically significant correlation between social
	463	deprivation status and age of the patient with longer travel distances. Patients from
	464	poorer sociodemographic background may be expected to travel further for RevKR.
	465	This highlights the additional care coordination and follow-up efforts that should
	466	accompany the widening reach of regional revision knee networks. It is reassuring
	467	that access to treatment for older patients is unaffected by travel distance. However,
	468	there may be patients who refused to travel to a specialist centre and opted for
34	469	treatment at their local centre.
35 36	470	
37 38	471	
39 40	472	
41 42	473	Meaning of the study: possible explanations and implications for clinicians and
43 44	474	policymakers
45 46	475	
47	476	The organisation and delivery of revision knee services in England has recently
48 49	477	undergone a substantial change and now such services are provided around
50 51	478	regional networks of care. This promises substantial advantages to the increasing
52 53	479	number of patients with problematic knee replacements in our ageing population who
54	480	will benefit from regional expertise.[8] However, it is unknown the impact of patients
55 56	481	residing farther from tertiary referral centres, particularly rural patients who may
57 58	482	encounter additional difficulties associated with greater travel distance. A recent

59 483 study following the outcomes of aortic surgery found that longer travel distances are 60

BMJ Open

2		
3 4	484	associated with inferior perioperative outcomes[13]. Similar associations have been
5 6 7 8	485	found in postoperative colorectal surgery patients [32]. As such our results are
	486	reassuring to policy makers and clinicians.
9 10	487	
11	488	Unanswered questions and future research
12	489	
14 15	490	There is a correity of ovidence evaluating the national percention of complex health
16	491	interventions auch as naturally models of some Depend work by Kurder et al bas
17 18	492	Interventions such as network models of care. Recent work by Kugler et al has
19 20	493	demonstrated the willingness of patients to travel further for better outcomes in the
20	494	context of total knee replacement surgery. [33] Nevertheless, patient perceptions of
22 23	495	travelling further for their treatment should be a focus for future research in the
24	496	context of revision knee patients, particularly as this is one of the top ten research
25 26	497	priorities identified by the James Lind Alliance priority setting partnership.[34]
27 28	498	
29 30 31 32	499	
33	500	Conclusion
34 35	501	
36	502	We did not observe an association in our study population between 30-day
37 38 39 40	503	readmission rates and increasing travel distances or times between a patient's home
	504	and their treating hospital in revision knee replacement. This paper is the first to
41	505	explore the relationship between travel distance and complex orthopaedic surgery
42 43	506	and informs some concerns regarding the creation of a centralised revision knee
44 45	507	network. This information is of utility to surgical providers and commissioners of
46	508	healthcare services. Furthermore, it can inform patient-led decision making and the
47 48	509	exploration of perceptions surrounding travelling for complex surgery. Although this
49 50	510	is the first assessment in complex orthopaedic surgery, a prospective analysis will be
51 52	511	undertaken as part of the ongoing auditing of revision knee networks in England.
53 54 55 56	512 513	
57 58	514	Supplementary material and figures
59 60	515	

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

BMJ Open

1

2 3	-16	
4	516	
5 6	517	Supplementary material S1 – Journey Time Statistics Reference Document
7	518 510	
8 9 10	520	Supplementary material S2 – OPCS-4 code criteria used for Hospital Episode
10	521	Statistics data extraction
12 12	522	
14	523	
15 16	524	See separate file named supplementary material S2
17	525	
18 19	526	
20	527	
21 22	500	Quertamentary material C2. Elever of a stight inclusion (such sizes
23	528	Supplementary material S3 – Flow of patient inclusion/exclusions
24 25	330	
26	531	-See attached file named Supplementary Material S3
27 28	522	
29 20	332	
30 31 32 33 34 35	533	Supplementary material S4 – R Code
	534 535	See attached file named Supplementary Material S4
	555	See allached hie hamed Supplementary Material S4
35 36	536	
37	537	
38 39	538	Supplementary material S5 –Scatterplot for imputed data: A comparison
39 40 41	539	between imputed values and observed values following multiple random
42 43	540	imputation. Imputed values in "blue", observed values in "grey". Imputation 0
44 45	541	on X axis refers to original dataset. Subsequent random imputations labelled 1
45 46	542	to 5 on x axis.
47 48	543	
49	544 545	
50 51	545 546	
52	547	
53 54	548	
55		
56 57		
58		
59 60		

BMJ Open

2 3		
4	549	Figure 1 -
5 6	550	
7 8	551	(Left) Scatterplot showing correlation between patient age and travel distance.
9 10	552	Red line represents linear regression trend. Spearman's rank correlation is
11 12	553	presented in chart.
13 14	554	
15 16	555	(Right) Scatterplot showing correlation between social deprivation and patient
17 18	556	travel distance. Red line represents linear regression trend. Spearman's rank
19 20 21 22 32 42 52 62 72 82 93 31 32 33 43 53 63 73 83 940 41 23 44 546 47 849 50 152 53 54 55 56 57 859 60	557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587	correlation is presented in chart.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1

2		
3 4	588	Figure 2 - Predicted probability of emergency readmission at 30 days by
5 6 7 8	589	straight line patient travel distance from hospital after RevKR
	590	A Fixed effects multivariable logistic regression model using 3 knots at 5%,
9 10	591	50% and 95% centiles of mean unit volume. 95% confidence intervals
11 12 13 14	592 593 594	represented by blue shaded line
15 16 17 18 19 20 21 22 23 24 25 26 27	595	
	596	
	597	
	598	
	599	
	600	Figure 3 - Predicted probability of prolonged length of inpatient stay at by
28 29	601	patient straight line travel distance from hospital after RevKR
30 31	602	A Fixed effects multivariable logistic regression model using 4 knots at 5%,
32	603	35%, 65% and 95% centiles of mean unit volume. 95% confidence intervals
33 34	604	represented by blue shaded line
35 36	605	
37 38	606	
39 40	607 608	
41 42	609	Figure 4 - Predicted probability of prolonged length of inpatient stay at by
43 44	610	patient driving distance from hospital after RevKR
45 46	611	A Fixed effects multivariable logistic regression model using 4 knots at 5%.
47 48	612	35%. 65% and 95% centiles of mean unit volume. 95% confidence intervals
49 50	613	represented by blue shaded line
51 52	614	
53 54	615 616	
55 56		
57 58		
59 60		
00		

1 2		
3 4	617	Figure 5 - Predicted probability of prolonged length of inpatient stay at by
5 6	618	patient driving time from hospital after RevKR
7 8	619	A Fixed effects multivariable logistic regression model using 4 knots at 5%,
9 10	620	35%, 65% and 95% centiles of mean unit volume. 95% confidence intervals
11 12	621	represented by blue shaded line
13 14	622 622	
15 16	023	
17 18		
19 20		
21 22		
23 24		
25 26		
27 28		
29 30		
31 32		
33 34		
35 36		
37 38		
39 40		
41 42		
43 44		
45 46		
47 48		
49 50		
51 52		
53 54		
55 56		
57 58		
59 60		

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1 2		
3	624	
4 5	625	
6 7	626	
8	627	
9 10 11	628	Contributorship
12 13	629	
14 15	630	
16	631	Alex Matthews: Conceptualisation, Methodology, Project Administration,
17 18	632	Investigation, Data Curation, Formal Analysis, Visualisation, Writing - original draft,
19 20	633	Writing - review and editing. This author is the guarantor and is responsible for the
21	634	content
22	635	
24 25	636	Jonathan P Evans: Conceptualisation, Supervision, Writing - review & editing
26 27	637	
28	638	Jonathan T Evans: Supervision, Writing - review and editing
29 30	639	
31 32	640	Sarah E Lamb: Conceptualisation, Supervision, Writing - review and editing
32 33 34	641	
34 35	642	Andrew Price: Conceptualisation, Supervision, Writing - review and editing
36 37	643	Le la
38 39	644	William Gray: Conceptualisation, Supervision, Methodology, Writing - review and
40	645	editing
41 42	646	
43 44	647	Tim Briggs: Supervision, Writing - review and editing
45	648	
40 47 48	649	Andrew Toms: Conceptualisation, Supervision, Writing - review and editing
49	650	
50 51	050	
52 53	651	
54	652	
55 56	653	
57	654	
58	655	
59 60	656 657	
	037	

8 Public and Patient Involvement statement

The study's chief investigator (AT) led the James Lind Alliance 'Revision Knee Replacement' priority setting partnership. This group of patients, carers and health care professionals identified the need to investigate the best way of organising revision knee replacement surgery to improve patient outcomes as one of their top 10 research questions. Patients were therefore directly involved in the development of the study's aims and objectives. The results of the study will be disseminated to the members of this group prior to publication.

8 Copyright/licence for publication Statement

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence.

676 Competing Interests

All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

686 Transparency Declaration

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

The lead author (the manuscript's guarantor) affirms that the manuscript is an

honest, accurate, and transparent account of the study being reported; that no

important aspects of the study have been omitted; and that any discrepancies from

the study as planned (and, if relevant, registered) have been explained.

Ethical Approval

The analysis and presentation of data follows current NHS England guidance for the use of HES data for research purposes and is anonymised to the level required by ISB1523 Anonymisation Standard for Publishing Health and Social Care Data. Ethical approval was not required.

Funding

No funding was obtained to carry out this study

Data Sharing

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

2		
3	719	1. Price AJ, Alvand A, Troelsen A, Katz JN, Hooper G, Gray A, et al. Knee replacement.
4	720	Lancet. 2018:392(10158):1672-82.
5	721	2. Sabah SA, Hedge FA, Abram SGE, Alvand A, Price AL, Hopewell S, Patient-reported
7	722	outcome measures following revision knee replacement: a review of PROM instrument
8	722	utilisation and measurement properties using the COSMIN checklist BMI Open
9	723	
10	724	2021,11(10).0040109.
11	725	3. Sabari S, Knight R, Alvanu A, Murray D, Petrou S, Beard D, Price A. No exponential
12	/26	rise in revision knee replacement surgery over the past 15 years: an analysis from the
13 14	727	National Joint Registry. Osteoarthritis and Cartilage. 2022;30(12):1670-9.
15	728	4. Rowland T, Lindberg-Larsen M, Santy-Tomlinson J, Jensen CM. A qualitative study of
16	729	patients' experiences before, during and after surgical treatment for periprosthetic knee
17	730	joint infection; "I assumed it had to be like that ". International Journal of Orthopaedic
18	731	and Trauma Nursing. 2023;48:100992.
19	732	5. Garbarino LJ, Gold PA, Sodhi N, Anis HK, Ehiorobo JO, Boraiah S, et al. The effect of
20	733	operative time on in-hospital length of stay in revision total knee arthroplasty. Ann Transl
21 22	734	Med. 2019;7(4):66.
22	735	6. Liodakis E. Bergeron SG. Zukor DJ. Huk OL. Epure LM. Antoniou J. Perioperative
24	736	Complications and Length of Stay After Revision Total Hip and Knee Arthroplasties: An
25	737	Analysis of the NSOIP Database The Journal of Arthroplasty 2015:30(11):1868-71
26	738	7 Briggs TWR A National Review of Adult Elective Orthonaedic Services in England
27	730	In: Time GIRE editor 2015
28	737	 Kalson NS Mathews IA Phillips IRA Baker PN Price AI Toms AD et al Revision
29 30	740	knoo replacement curgery in the NHS: A PASK curgical practice guideline. The Knoo
31	741	
32	742	2021;29:353-64.
33	/43	9. Yapp LZ, Walmsley PJ, Woran W, Clarke JV, Simpson AHRW, Scott CEH. The effect of
34	/44	hospital case volume on re-revision following revision total knee arthroplasty. J Bone Joint
35	745	Surg Br. 2021;103-B(4):602-9.
30 27	746	10. Halder AM, Gehrke T, Günster C, Heller KD, Leicht H, Malzahn J, et al. Low Hospital
38	747	Volume Increases Re-Revision Rate Following Aseptic Revision Total Knee Arthroplasty: An
39	748	Analysis of 23,644 Cases. J Arthroplasty. 2020;35(4):1054-9.
40	749	11. Jeschke E, Gehrke T, Günster C, Heller KD, Leicht H, Malzahn J, et al. Low Hospital
41	750	Volume Increases Revision Rate and Mortality Following Revision Total Hip Arthroplasty: An
42	751	Analysis of 17,773 Cases. J Arthroplasty. 2019;34(9):2045-50.
43	752	12. Bloch BV, Matar HE, Berber R, Gray WK, Briggs TWR, James PJ, Manktelow ARJ. The
44 45	753	impact of a revision arthroplasty network on patient outcomes. J Bone Joint Surg Br.
46	754	2023:105-b(6):641-8.
47	755	13. Feldman 7M. Zheng X. Mao I. Sumpio BI. Mohebali I. Chang DC. et al. Longer patient
48	756	travel distance is associated with increased non-index readmission after complex aortic
49	757	surgery Journal of Vascular Surgery 2023:77(6):1607-17 e7
50	759	14 Soundars BS Fornandos Taylor S Kind AL Engelbort TL Greenborg CC Smith MA et
51 52	750	14. Sounders NS, Fernandes-Taylor S, Kind AJ, Engelbert TE, Greenberg CC, Smith MA, et
52 53	739	al. Renospitalization to primary versus different facilities following apdominal aortic
54	/00	aneurysm repair. J vasc Surg. 2014;59(6):1502-10, 10.01-2.
55	/61	15. Sinna S, Peach G, Poloniecki JD, Thompson MIM, Holt PJ. Studies using English
56	/62	administrative data (Hospital Episode Statistics) to assess health-care outcomessystematic
57	763	review and recommendations for reporting. Eur J Public Health. 2013;23(1):86-92.
58 50		
60 59		
00		

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

16. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344-9. 17. NHS Digital. Hospital Episode Statistics(HES) Analysis Guide. London, UK: NHS Digital; 2018. 18. Information Standard Board for Health and Social Care. Anonymisation Standard for Publishing Health and Social Care Data Specification. London, UK: NHS Digital; 2013. NHS Digital. Delivering linked datasets to support health and care delivery and 19. research. London, UK: NHS Digital; 2018. 20. Department for Transport. Journey Time Statistics, notes and definitions. London, UK: Department for Transport; 2019. Kalson NS, Mathews JA, Miles J, Bloch BV, Price AJ, Phillips JRA, et al. Provision of 21. revision knee surgery and calculation of the effect of a network service reconfiguration: An analysis from the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. Knee. 2020;27(5):1593-600. 22. Deere K, Whitehouse MR, Kunutsor SK, Sayers A, Price AJ, Mason J, Blom AW. How long do revised and multiply revised knee replacements last? An analysis of the National Joint Registry. Lancet Rheumatol. 2021;3(6):e438-e46. 23. Kamath CC, O'Byrne TJ, Lewallen DG, Berry DJ, Maradit Kremers H. Neighborhood-Level Socioeconomic Deprivation, Rurality, and Long-Term Outcomes of Patients Undergoing Total Joint Arthroplasty: Analysis from a Large, Tertiary Care Hospital. Mayo Clin Proc Innov Qual Outcomes. 2022;6(4):337-46. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple 24. imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Medical Research Methodology. 2017;17(1):162. 25. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Ann Transl Med. 2016;4(2):30. 26. Sabah SA, Knight R, Alvand A, Palmer AJR, Middleton R, Abram SGF, et al. Patient-Relevant Outcomes Following First Revision Total Knee Arthroplasty, by Diagnosis: An Analysis of Implant Survivorship, Mortality, Serious Medical Complications, and Patient-Reported Outcome Measures Utilizing the National Joint Registry Data Set. J Bone Joint Surg Am. 2023;105(20):1611-21. Hardy F, Heyl J, Tucker K, Hopper A, Marchã MJ, Briggs TWR, et al. Data consistency 27. in the English Hospital Episodes Statistics database. BMJ Health Care Inform. 2022;29(1). 28. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021;374:n2061. Farrow L, Kennedy JW, Yapp L, Harding T, Walmsley P. Provision of revision knee 29. arthroplasty services across Scotland: A national audit. Knee. 2023;42:312-9. Paterson C, Baarts C, Launsø L, Verhoef MJ. Evaluating complex health interventions: 30. a critical analysis of the 'outcomes' concept. BMC Complement Altern Med. 2009;9:18. Jackson KL, Glasgow RE, Mone MC, Sheng X, Mulvihill SJ, Scaife CL. Does travel 31. distance influence length of stay in elective pancreatic surgery? HPB (Oxford). 2014;16(6):543-9. 32. Cloyd JM, Huang L, Ma Y, Rhoads KF. Predictors of readmission to non-index hospitals after colorectal surgery. The American Journal of Surgery. 2017;213(1):18-23.

1 2 3	810	33. Kugler CM. De Santis KK. Rombey T. Goossen K. Breuing I. Könsgen N. et al.
4 5	811 812	Perspective of potential patients on the hospital volume-outcome relationship and the
6 7	812 813	interview study, BMC Health Serv Res, 2021-21(1):633
8	814	34. Mathews JA, Kalson NS, Tarrant PM, Toms AD. Top ten research priorities for
9 10	815	problematic knee arthroplasty. Bone Joint J. 2020;102-b(9):1176-82.
11	816	
12		
13 14		
15		
16 17		
18		
19 20		
21		
22 23		
24		
25 26		
27		
28 29		
30		
31 32		
33		
34 35		
36		
37 38		
39		
40 41		
42		
43 44		
45		
46 47		
48		
49 50		
51		
52 53		
54		
55 56		
57		
58 59		
60		







360x119mm (300 x 300 DPI)

30 35 40 45 Travel Distance (Miles)

Figure 2

87x59mm (300 x 300 DPI)

50

55

60

65

70

75

20

25

30



BMJ Open: first published as 10.1136/bmjopen-2024-085201 on 6 May 2025. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.



Figure 3 76x59mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



Figure 4

BMJ Open: first published as 10.1136/bmjopen-2024-085201 on 6 May 2025. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

78x60mm (300 x 300 DPI)









72x60mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Journey Time Statistics: Notes and Definitions

About this release

This publication supports the latest statistics on journey times.

In this publication

Overviewp1
Access to key services
p4
Connectivityp7
Data sourcesp9
Outputsp18
Strengths and weaknesses
p19

Further information

Public enquiries 020 7944 3077 vehicles.stats@dft.gov.uk Media enquiries 020 7944 3066

Overview

This note provides information on the methodology used, the source data and definitions of key terms for calculating Journey Time Statistics.

These annual statistics were first published in December 2015 for the year 2014 and have been developed from the earlier Accessibility Statistics published for 2007 to 2013.

The Journey Time Statistics produced by DfT consists of theoretical journey times calculated by modelling journeys between known sets of origins and destinations. It uses information on the road network, traffic speeds and public transport timetables in England.

The relevant Journey Time Statistics calculation is varied for origins and destination to meet a variety of needs. Two sets of analysis are published: Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

- Access to key services; and
- Connectivity

Origin indicators

These indicators measure the number of different services in a particular area that users can reach within a given time.

Destination indicators

These indicators measure the proportion of users that can access a service within a certain time.

The 'user' populations for each service in the destination indicators are:

Employment	16-74 year olds		
Primary schools	5-10 year olds		
Secondary schools	11-15 year olds		
Further education	16-19 year olds		
All other services	All households		

Journey Time Statistics: Notes and Definitions - Page 1

Key services

- Employment centres: Data used are the number of jobs in a Lower Super Output Area (LSOA). The data tables include results for employment centres of 3 different sizes (100-499 jobs, 500-4,999 jobs and at least 5,000 jobs). For the key services average, the 500-4,999 jobs definition is used for employment.
- <u>Education</u>: Locations of all open Primary schools, Secondary schools, Further Education and Sixth Form Colleges.
- General Practice (GP) surgeries: For 2017 based on the Patients Registered at a GP Practice dataset released by NHS Digital – previously this was based on a filtered dataset of NHS prescribers released by NHS Digital.
- Hospitals: Based on hospitals that are registered with the Care Quality Commission (CQC) and are managed by Acute Trusts.
- ► <u>Food stores</u>: Locations of grocery, supermarkets or convenience stores.
- Town centres: Locations of Town centres using a central focal point for the town mapped to the nearest road.

Geography

Local authorities

In some parts of England there are two tiers of local authorities, and in others a single unitary authority. Statistics have been calculated for both types of authority - around 360 in all. These vary considerably in size, from a population of a few tens of thousands to over a million.

Lower Layer Super Output Areas (LSOA)

LSOAs are small areas designed to be of a similar population size, with an average of approximately 1,500 residents or 650 households. There are 32,844 Lower-layer Super Output Areas (LSOAs) in England. They were determined by the Office for National Statistics for the reporting of small area statistics and are derived from the 2011 Census.

Urban and rural definitions

This report uses the Defra Rural-Urban Classification, based on 2011 Census Output Areas. The Rural-Urban Classification defines areas as rural if they fall outside of settlements with more than 10,000 resident population. See <u>Defra's Definitions and Local Authority Classification</u> for more details.

Journey time calculations

The journey time calculations are carried out using a commercially available software package 2 3 called TRACC, owned by Basemap. TRACC is a desktop application that uses public transport and 4 highways data to create journey times from origins to destinations. It uses timetable information 5 6 showing both arrival and departure times at stops from public transport services against a specific 7 time/day period. Highways information from road networks are used to fill the gaps between public 8 9 transport services by creating a linear network that connects the origins, destinations and stops 10 together. This provides a fully routable network of nodes and lines which is saved on file as a graph 11 12 network. The graph network has various constraints which can be altered to suit the user need 13 such as distance travelled, interchange delays on public transport and stopping limitations on road 14 15 networks. The TRACC software then gueries the graph network with origin and destination co-16 17 ordinates and uses the Dijkstra shortest path algorithm to route between these points. This is an 18 algorithm for finding the shortest distance for travel between the graph networks. 19

20 For a public transport journey, the journey time produced includes all walking elements of the 21 journey, i.e. the walk from the origin of the journey to the road, from the road to public transport 22 23 stops, any interchange of public transport using the road and then from the final stop to the 24 destination via the road, and finally from the nearest point on the road network to the destination. 25 26 The journey assumes arrival at the first stop one minute before the initial departure, with any 27 subsequent interchange waiting times included as part of the final journey time. 28 29

Car, cycle or walk only journeys are similar except that once the road network is reached the
 journey proceeds link by link along the road network at speeds governed by data held in the model.
 These are specific to the mode, the road type, and in some cases the individual road link.

The 10 shortest journey times from each origin (i.e. Output Area) are calculated for each destination type. For the public transport / walking mode these consist of the 10 shortest journey times by either walking or public transport, after applying a 5 minute penalty for any journeys using public transport (to represent travellers arriving slightly early at the first stop).

The journey times are representative of the 'morning peak'. This is made explicit for public transport / walking by requiring the journey to be completed between 7 and 10am, and for car journeys by using average traffic speeds for between 7 and 10am. For the cycle mode no actual speed data are available. The cycle speeds used are default assumptions, and are not based on a particular time of day.

Access to key services

The Access to Services analysis applies the Journey Times methodology to origins consisting of residential neighbourhoods and destinations consisting of centres of employment and a range of key local services. Journey times are calculated for three modes of transport: public transport; driving; and cycling. These journey times are then used to generate further indicators, as described in the Outputs Section.

The Access to Services calculation process and the coverage of the data set are very similar to those of the Accessibility Statistics from which they were developed. However, the calculation algorithm and a number of other features of the design are different, so the results are not directly comparable.

The statistics are designed to represent as much as possible the situation on a Tuesday in October of the year to which they relate. Data for the second week of October are used in the analysis, since this provides a fairly typical week, unaffected by major national holidays, school holidays or other seasonal effects. The origins, destinations and public transport timetables used are as far as possible for this date. The traffic data are averages for the preceding 12 months up to and including August. The road networks are those current at the start of the traffic data year.

Outline of access to services calculation process

Origins 171,372 Output Areas (OA) (Census geography)

Destinations Employment locations (3 sizes) Education (Primary schools, Secondary Schools, Further Education colleges) Health (GPs, Hospitals) Food stores Town centres

> **Transport data Bus/rail timetables** Road network Average road speeds

Travel time calculation Using TRACC software, similar to running millions of journey planner queries



Output data

Travel times from each of 32,844 Lower Super Output Areas (LSOA) to nearest 10 of each destination

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Public transport / walk x1 time period AM peak

x3 modes Cycle Car

Model parameters and assumptions

General parameters

Maximum journey time of 2 hours.

Maximum journey distance of 100km.

Walking

60

These apply to both:

walking between origin / destination and the transport networks at both ends of a journey by

Journey Time Statistics: Notes and Definitions - Page 4

2 3

4

5 6

7

8

any mode;

walk only journeys as part of the public transport / walk mode.

Maximum straight line distance between origin / destination and road network of **2km**. The algorithm will always use nearest point on network. For cycle or car modes, travel by cycle or car begins from this point. For public transport/walk, traveller walks along road network to the most suitable public transport stop, or direct to the destination if this is quicker.

- ⁹ Walking speed on road/path network of **4.8km/h.**
- ¹¹ Walking speed off road/path network of **4.0km/h.**

13 **Public transport**

Interval within which door-to-door journey must be completed (required for timetable selection) is
 7am to 10am on a Tuesday.

¹⁷Maximum walk distance of **3km** - this applies to walks from origin to first public transport stop, from ¹⁹last stop to destination, and also walking directly from origin to destination without using public ²⁰transport at all.

Maximum number of potential first public transport stops considered in routing algorithm is 100 (starting with the closest to origin).

Allowance for catching first public transport service is **5 minutes** - added to any journey that involves boarding one or more public transport services.

Public transport speed – this is provided implicitly by the timetable information.

Interchange time of 5 minutes (minimum interval allowed between arriving at a stop and catching
 another service).

 $^{33}_{34}$ Maximum straight line distance between public transport interchanges of **500m**.

Stop clustering at 150m – groups together public transport stops within this distance of one another
 to speed up processing. The individual timetables for each service are retained.

Cycling speeds

38 39 40

58 59 60

Road Type	Speed
Motorway	0.0 km/h
Urban Motorway	0.0 km/h
A road	16.0 km/h
B road	16.0 km/h
Minor road	16.0 km/h
Local street	16.0 km/h
Private road – restricted access	4.8 km/h
Private road – public access	16.0 km/h
Pedestrian street	4.8 km/h
Alley	4.8 km/h

Parking time of **5 minutes** - added to all cycle journeys.

Car speeds

-				
Type of road	2014	2015	2016	2017
		Default spe	eeds (km/h)	
Motorway	79.5	77.0	77.5	77.6
Urban Motorway	79.5	77.0	77.5	77.6
A road	42.7	43.7	43.3	43.2
B road	41.6	43.0	42.2	41.9
Minor road	36.8	37.5	36.8	36.3
Local street	19.2	17.8	18.8	18.3
Private road – restricted access	17.0	16.7	16.2	15.3
Private road – public access	14.8	15.2	15.1	13.6
Pedestrian street	0.0	0.0	0.0	0.0
Alley	0.0	0.0	0.0	0.0

Car speeds are calculated for specific links where more than 200 records exist otherwise the default speeds are used. Minimum journey time for a journey that uses a car is **5 minutes**.

Time at junctions

Road normalisation is used for all modes of transport which converts each road link to a straight line to speed up processing. The true link length is retained for accurate speed/time calculations, but there could be a small effect on the calculation of shortest distance from the road network to destination points. Effect for origins is minimal due to origins being constrained to road nodes.

Connectivity

These experimental analyses are intended to apply the Journey Times methodology to a range of more strategic or economically significant destinations than the primarily local services covered by the Access to Services analyses; including airports and railway stations. The principle difference in the Connectivity approach from that of the Access to Services analyses is that journey times are calculated, as far as possible, to all accessible locations, rather than to just the nearest 10 examples. This tends to result in a much larger data set being generated. In some cases a longer maximum journey time may be allowed although this may depend on what is considered reasonable for the type of destination. Given these factors, a less detailed origin data set may be used than for Access to Services. This is both necessary, to limit the size of the data set, and acceptable where the typical journey lengths are longer.

The first connectivity analyses published using the new Journey Time methods were released in Journey Time Statistics 2015, published in April 2017, for two destination sets – airports and rail stations. These analyses using the Journey Times methods superseded two earlier Connectivity Statistics reports published in 2014 and 2015 based on the old accessibility statistics methods, in the same way that the new Access to Services analyses have replaced the earlier Accessibility Statistics. Again, the connectivity results produced using the old and new methods are not directly comparable.

Outline of Connectivity calculation



Model parameters and assumptions

Origins	Population weighted centroids (the central
	point) of 32,844 English LSOAs as specified in
	the 2011 Census geography. These points were
	then constrained to the nearest road node, as
	for Access to Services method.
Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

BV	AJ Open Page
Journey Time Calculation	As for Access to Services, for public transport
	/ walking and car modes only, except that a
	maximum journey time of 240 minutes and
	maximum straight line distance of 400km is
	allowed.
Outputs	Generally similar to Access to Services,
•	with different journey time classifications as
	appropriate Journey time results to specific
	destinations are included – this is the key
	difference in the Connectivity analyses
	(Average inverse) times' and (nearest)
	Average journey times and nearest
	destinations should be used with caution.
	The average journey times exclude results
	for areas with no available connection under
	240 minutes, which may become significant
	in remote areas and for destinations are a
	great distance from the origin. The 'nearest'
	destination is the destination with the shortest
	average journey time across the whole area
	appointed which will be relatively large in the
	soon of local outbarity lovel use I've
	case of local authority level results.

review only

http://bmjopen.bmj.com/site/about/guidelii

Data sources

Origins

The origins used for all Access to Services calculations are the 171,372 English Output Areas (OA) as specified in the 2011 Census geography.

To provide the actual journey start point in each OA, the population weighted centroid of the OA was shifted to the nearest node (i.e. junction) on the road network. This was to avoid biasing the journey time results where the centroid of the OA was a long way from a road. In fact it is rare for an OA centroid to be more than about 100 metres from a road – only a tiny handful of OA in remote areas have centroids as much as 1km from a road. The OA centroids have been shifted onto the nearest road node rather than the nearest point on a road in order to reduce issues arising from normalising the road network.

Origin	Data source for the origin points
All	Data: Population centroid of each Output Area in
	2011.
	Source: ONS 2011 Census Boundaries.
	Further information: <u>http://geoportal.statistics.gov.ul</u>

Destinations

The destinations used consist of three different sizes of employment centre and the locations of seven other types of key local service. For each of these key services a nationally consistent data set has been identified or derived – further information on these is provided in this section.

Each destination is located by a 6-figure National Grid reference. For the employment destinations this is taken to be the population weighted centroid of the LSOA.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

Destination	Number of locations			
	2014	2015	2016	2017
Employment centres (small)	16,465	16,625	16,930	17,194
Employment centres (medium)	9,235	9,460	9,707	10,241
Employment centres (large)	645	676	719	785
Primary schools	16,463	16,484	16,655	16,927
Secondary schools	3,365	3,376	3,381	3,174
Further education colleges	2,624	2,606	2,418	2,304
GPs	9,257	11,167	9,128	7,353
Hospitals	296	278	278	277
Food stores	19,549	19,746	21,665	20,987
Town centres	1,211	1,211	1,211	1,211

The data source for GP surgeries was reviewed and replaced for 2017.

Access to key services

Destinations 2017	Data source for the locations	Data source for users of the
	of the service	service
Employment	Data: Number of jobs available in a LSOA in the year before the calculation year.	Data: Number of 16-74 year olds in each output area.
	Source: ONS Business Register Employment Survey.	Source: ONS mid-year population estimates for calculation year.
	Further information: <u>https://</u> www.nomisweb.co.uk/default. asp	Further information: ONS mid-year population estimates: http://www.ons.gov.uk/ons/ taxonomy/index.
Primary schools	Data: Location of all open primary schools in September of calculation year.	Data: Number of 5-10 year old in each output area.
	Source: The Department for Education (DfE) Edubase.	Source: ONS mid-year population estimates for calculation year.
	Further information: <u>https://get-information-schools.service.</u> gov.uk/	Further information: ONS mid-year population estimates: http://www.ons.gov.uk/ons/ taxonomy/index.
Secondary schools	Data: Location of all open secondary schools in September of calculation year.	Data: Number of 11-15 year olds in schools in each output area.
	Source: DfE Edubase.	Source: ONS mid-year population estimates for calculation year.
	Further information: <u>https://get-information-schools.service.</u> <u>gov.uk/</u>	Further information: ONS mid-year population estimates: http://www.ons.gov.uk/ons/ taxonomy/index.
Further education colleges	Data: Location of all open further education and sixth form colleges/school sixth form in September of calculation year.	Data: Number of 16-19 year olds in each output area.
	Source: DfE Edubase.	Source: ONS mid-year population estimates for calculation year.
	Further information: <u>https://</u> get-information-schools. service.gov.uk/	Further information: ONS mid-year population estimates: http://www.ons.gov.uk/ons/ taxonomy/index.

Journey Time Statistics: Notes and Definitions - Page 10

For peer review only - http://bmjopen.bmj.com/site/about/guidelin

Destinations 2017	Data source for the locations	Data source for users of th
	of the service	service
GPs	Data: Locations of GP	Data: Number of households
	surgeries with registered	each output area.
	patients in October of	
	calculation year	
	Source: NHS Digital table of	Source: 2011 Census +
	Registered natients at GP	Local Authority (LA) undates
	practices	from the Ministry of Housing
	practices	Communities & Local
		Covernment (MHCL C) mid
		Government (MinCLG) Ind-
	Further information: <u>https://</u>	Further information: 2011
	digital.nns.uk/data-and-	Census: <u>http://www.nomiswe</u>
	Information/publications/	<u>co.uk/census/2011</u>
	statistical/patients-registered-	MHCLG mid-year household
	at-a-gp-practice	projections: https://www.gov
		uk/government/statistical-da
		sets/live-tables-on-househol
		projections
Hospitals	Data: Location of hospitals.	Data: Number of households
		each output area.
	Source: Care Quality	Source: 2011 Census + LA
	Commission - Directory of	updates from MHCLG mid-
	places that provide care.	year household projections of
		calculation year.
	Further information: http://www.	Further information: 2011
	cac.org.uk/content/how-get-	Census: http://www.nomiswe
	and-re-use-coc-information-	co.uk/census/2011
	and-data	
		MHCLG mid-year nousenoid
		projections: <u>nttps://www.gov</u>
		<u>uк/government/statistical-da</u>
		sets/live-tables-on-househol
		projections

http://bmjopen.bmj.com/site/about/guid

review only

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Destinations 2017	Data source for the locations	Data source for users of th
	of the service	service
Food stores	Data: Location of grocery/	Data: Number of households
	supermarkets or convenience	each output area.
	stores in October of calculation	
	year.	
	Source: The Local Data	Source: 2011 Census + LA
	Company	updates from MHCLG mid-
		vear household projections of
		calculation year
	Further information: https://	Further information: 2011
	www.localdatacompany.com/	Census: http://www.nomiswe
		co.uk/consus/2011
		MHCLG mid-year household
		projections: <u>https://www.gov.</u>
	4	uk/government/statistical-dat
		sets/live-tables-on-househole
		projections
Town centres	Data: Location of town centres	Data: Number of households
	in 2004.	each output area.
	Source: MHCLG Town Centre	Source: 2011 Census + LA
	and retail planning statistics for	updates from MHCLG mid-
	England and Wales.	year household projections of
		calculation vear.
	Further information: https://	Further information: 2011
	data.gov.uk/dataset/	Census: http://www.nomiswe
	ed07b21f-0a33-49e2-9578-	co.uk/census/2011
	83ccbc6a20db/english-town-	
	centres-2004	MHCLG mid-year household
		projections: <u>https://www.gov.</u>
	U	uk/government/statistical-dat
		sets/live-tables-on-household
		projections

http://bmjopen.bmj.com/

/site/about/gu

2 3

4

5

6 7

8

9

48

57

58

GP destination data

The GP surgery destinations used from 2014 to 2016 are based on the list of practices maintained by the Organisational Data Service of the Health & Social Care Information Centre, and published at https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-

related-data. This was supplemented with information on branch surgeries from the same source. Grid references were derived from the postcode using the Office for National Statistics (ONS)

Postcode Address File. Practices with identical postcodes were taken to be duplicates or colocated, and all additional records after the first were removed. 10

11 From 2017, the list of GP locations is taken from the NHS Digital publication of Registered patients 12 at GP practices for October of the calculation year. This had the effect of reducing the number 13 14 of locations in the dataset, but removed the need for manual adjustments and produces a more 15 stable list defined as GP practices with registered patients. Grid references were derived from the 16 postcode using the Office for National Statistics (ONS) Postcode Address File. 17

18 Hospital destination data 19

20 The starting point for hospital sites is the Care Quality Commission's (CQC) list of 'active locations' 21 dataset, which is thought to be the most-up-to date and freely available source of data on individual 22 23 National Health Service (NHS) and social care 'sites' or hospitals. A criteria was developed in 24 consultation with the Department of Health to reduce the list down to capture only the key hospitals. 25 The following have been removed and individual records have been inspected to remove further 26 27 examples of these cases and for any duplicates: 28

- 29 care home records; •
- 30 non-NHS providers; 31
- 32 sites not associated with acute providers; ٠
- 33 any remaining sites that are associated with Specialist Trusts (usually single speciality Trusts or • 34 35 Sites):
- 36 records where it is evident from the name that the record is not a hospital (e.g. headquarters, 37 specialist units.) 38

39 This gave a final list of 278 hospitals in 2017 run by Acute (non-specialist) Trusts. As well as 40 covering all general hospitals this will still include some with a largely or entirely community or 41 42 rehabilitation role, where these happen to be managed by an Acute Trust. It was considered on 43 balance better to leave these in the list, rather than risk adding further subjectivity to the selection. 44 Whilst not perfect, it is considered that the resulting list is a significant improvement on that used 45 46 previously. 47

Steps taken to produce hospital data set

49 Remove records where Care Home = Y 50

51 Remove records where Provider ID begins 1-52

53 Keep records where **Benchmark Group** is Care Home or **Cluster Group** is Acute 54

Filter the trust site locations by name to remove obvious non-hospital sites. Key words 55 56

used for this process are: birth, dental, house, clinic, grange, lodge, infirmary, health, community, unit, surgery, centre

59 Manual review of remaining locations 60

Journey Time Statistics: Notes and Definitions - Page 13

Employment destination data

The employment centres are defined by the number of jobs existing in each English LSOA, taken from the Business Register Employment Survey. Large Employment Centres are defined as those with 5,000 or more jobs, Medium Employment Centres as those with 500 or more jobs, up to 4,999 and Small Employment Centres as those with 100 or more jobs, up to 499.

Data are downloaded from the Nomis website; although LSOA level BRES data has safeguarded
 access, access can be requested through the site. The chosen data download options are
 LSOA2011 geography, date as calculation year, variable as employment status where the value is
 employed, and the measure chosen is a count.

For the 2016 destination set, the BRES changed from 2001 census geography to 2011 census geography. The majority of LSOA boundaries are unchanged between these datasets, but some have been merged or split. Therefore the employment destination indicators are not strictly comparable between 2015 and 2016 Journey Time statistics. See <u>https://www.ons.gov.uk/</u> <u>methodology/geography/ukgeographies/censusgeography</u> for further information.

Education destination data

The education destination datasets are taken from the Department for Education database of educational establishments. The database was filtered to remove those establishments that were not open during the school year starting in September of the calculation year. Further filters were applied to remove special educational establishments, boarding schools and selective schools, and then to select schools at each phase of education for primary and secondary schools and further educational establishments. The following table lists the filters used.

iez oni

Phase of	Code Variable	Variable	Selec	ted codes and values
Education				
All Schools	OpenDate			30/08/17 or earlier; NUL
	CloseDate	,		30/08/18 or later; NULL
	TypeOfEstablishment_	TypeOfEstablishment	1	Community school
			2	Voluntary aided schoo
			3	Voluntary controlled
			-	school
			5	Foundation school
			6	City technology collect
			12	Foundation special
				school
			18	Further education
			28	Academy sponsor led
			29	Higher education
		N.		institutions
			31	Sixth form centres
			32	Special post 16
				institution
			34	Academy converter
			35	Free schools
			36	Free schools special
		14	39	Free schools 16 to 19
			40	University technical
		C		college
			41	Studio schools
			45	Academy 16-19
				converter
			46	Academy 16 to 19
				sponsor led
	Boarders_Code_	Boarders	0	Not applicable
			1	No boarders
			9	NULL
	AdmissionsPolicy_Code_	AdmissionsPolicy	0	Not applicable
			4	Non-selective
			9	NULL
	PhaseOfEducation Code	PhaseOfEducation	2	Primary
Primary				.

Journey Time Statistics: Notes and Definitions - Page 15

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Phase of	Code Variable	Variable	Selec	ted codes and values
Secondary	PhaseOfEducation Code		0	Not applicable
		ThaseOrEducation		
schools			4	Secondary
			5	Middle deemed secondary
			7	All through
	Statutory High age		>=16	
	Statutory Low age		< 16	
FE	PhaseOfEducation_Code_	PhaseOfEducation	4	Secondary
			5	Middle deemed secondary
			6	16 plus
			7	All through
	Statutory High age	·	>16	
	OfficialSixthForm_Code_	OfficialSixthForm	0	Not applicable
			1	Has a sixth form
			9	NULL
		OR		
FE	EstablishmentTypeGroup	EstablishmentTypeGroup	1	Colleges
	code_			

Food Stores destination data

The food stores destination dataset is purchased from <u>The Local Data Company</u> and includes all branches of multiple food store chains. Although some data are available for independent food stores, this only exists within town centres and so has not been included.

Connectivity

Destinations	Data source for the locations	Data source for users of the
Destinations Airports	Data source for the locations of the service Data: Location of GB airports excluding highlands and islands of Scotland Source: National Public Transport Access Nodes Further information: https:// data.gov.uk/dataset/ ff93ffc1-6656-47d8-9155- 85ea0b8f2251/national-public- transport-access-nodes-naptan	Data source for users of the service Data: Number of households in each output area. Source: 2011 Census + LA updates from MHCLG mid- year household projections of calculation year. Further information: 2011 Census: http://www.nomisweb. co.uk/census/2011 MHCLG mid-year household projections: https://www.goy.
		uk/government/statistical-data- sets/live-tables-on-household- projections

Destinations	Data source for the locations of the service	Data source for users of the service
Railway stations	Data: Location of larger (category A, B and C1) rail stations in GB Source: Network rail classification	Data: Number of households in each output area. Source: 2011 Census + LA updates from MHCLG mid- year household projections of
	Further information: <u>http://webarchive.</u> <u>nationalarchives.gov.</u> <u>uk/20101007153226/</u> http://www.dft.gov.uk/pgr/	Calculation year. Further information: 2011 Census: <u>http://www.nomisweb.</u> <u>co.uk/census/2011</u> MHCLG mid-year bousehold
	rail/passenger/stations/ betterrailstations/ http://archive.nr.co.uk/	projections: <u>https://www.gov.</u> uk/government/statistical-data- sets/live-tables-on-household-
	browse%20documents/ rus%20documents/route%20 utilisation%20strategies/	projections
	network/working%20 group%202%20-%20stations/ networkrusstations.pdf	

Transport network data

Travellers moved between their original and their destination via one or more of the following transport networks, depending on the mode of transport being modelled. For all modes, travellers will probably also need to walk between their origin / destination and the transport network. For some short journeys, it may be quicker for travellers to walk directly to their destination, rather than using public transport at all – this is why public transport / walking results are modelled as a combined mode.

Public transport

National public transport timetable data are publically available. Data for bus, local coach and other local transport services (e.g. light rail, metro, and ferry) are captured in the Traveline National Data Set (TNDS), rail timetable data are published by the Association of Train Operating Companies (ATOC), and national coach services in the National Coach Data Set (NCDS).

Walk

The walking network is represented by the road and urban path elements of the Integrated Transport Network produced by the Ordnance Survey.

Cycle

The cycling network is represented by the road network including cycle paths and bridleways from the Integrated Transport Network. Cycle journeys are also allowed to use footpaths at walking pace.

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

Car

The car network is represented by the road component of the Integrated Transport Network.

Data on actual vehicle speeds on each road network link (generally the stretch of road between 2 nodes, or junctions) is obtained from Trafficmaster Satnav devices and are used to estimate car speeds. These data are used to calculate annual average traffic speeds on each link of the road network (by direction if the link is bi-directional). These are used as the link speeds for cars in the modelling. Where the Trafficmaster sample for an individual link is too small, national averages of the same data for the particular road type are used instead. This is an innovation from 2014. Previously the sample was too small and the model reverted to default assumptions for car speed based on road type which were much higher than the Trafficmaster averages, resulting in some inconsistency in the model.

Outputs

The journey time results are used to create the following indicators for publication:

Indicator		Description
Minimum journey time		The shortest of the ten journey time results.
Origin indicators	6	Four measures, the number of destinations (up
		to the maximum of 10) that can be reached
		from a given origin within 15, 30, 45 and 60
		minutes.
Destination indicators		Four measures, the percentages of service
		users within the given geographical area who
		can access at least one service location within
		15, 30, 45 and 60 minutes.

Each of these indicators is calculated for each mode and each destination type, and at a number of geographical scales as follows:

England

Region

► Local Authorities, including London Boroughs, Metropolitan districts, Unitary authorities, Counties and non-Metropolitan districts, also Inner and Outer London and former Metropolitan counties

- ► 2011 Lower layer Super Output Area
- ► 2011 Defra Rural/Urban Classification

The indicators for each geography are calculated as population weighted averages. In other words, the average minimum journey time for an area, B, is:

mjt(B)= ∑(i=1)^n(mjt(OAi)×pop(OA_i))/pop(B)

where mjt(B) is the minimum journey time in area B, mjt(OAi) is the minimum journey time of the ith of n output areas making up area B, and pop(B) and pop(OAi) are the user populations resident in area B and output area i respectively.

The service user populations used in the above weighting, and in the destination indicators, depend on the destination type, as follows:

Destination type	Service user population basis
Employment centres	Resident population of working age (16-74
	years)
Primary schools	Population aged 5-10
Secondary schools	Population aged 11-15
Further education colleges	Population aged 16-19
GPs, hospitals, food stores, town centres	Number of households
Average key services	Resident population of working age (16-74
	years)

Strengths and Weaknesses

In using the data, the following points should be kept in mind:

All journey times are compiled on a consistent basis across the country.

► The statistics are based on the calculation of theoretical journey times, they are not based on real journeys. They are however based on actual public transport times, and average traffic speeds on the road network.

Although the statistics are calculated to a high level of geographical detail, some assumptions and simplifications are necessary in the modelling (for example assigning the start point of journeys to a single point in each Output Area, road speeds, interchange times for public transport).

► For 2016 we have used the 2015 BRES data to designate Lower Super Output Areas as employment centres. The 2015 BRES is the first year to use LSOAs based on the 2011 census, and although the majority of these are an exact match to the 2001 LSOAs, there are some that were merged, split or had other boundary changes. For these areas journey times from earlier years are not comparable to the 2016 journey times. This effect is more pronounced for large employment centres, as there are fewer destinations to route to.

► For particular areas, local authorities and other experts may have more detailed information allowing them to produce more accurate or detailed models of the local situation.

Demand responsive services (e.g. bus services which have to be booked) are only included to the extent that they can be plausibly modelled, in the Traveline National Data Set.

► Since new journey calculation software was adopted for 2014, along with a significant number of other changes to the methodology, from 2014 results are not directly comparable with those for earlier years.

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

3
4
5
6
7
/ 0
8
9
10
11
12
13
14
15
16
17
17
18
19
20
21
22
23
24
25
26
20
27
28
29
30
31
32
33
34
35
36
20
3/
38
39
40
41
42
43
44
15
75 76
40
4/
48
49
50
51
52
53
54
55
55
56
57
58
59
60

Supplementary material S1 – OPCS-4	code criteria	used for	Hospital	Episode
Statistics data extraction				

Code	Code description
OPCS-4 cc	odes for knee revision procedures
O180	Conversion from previous hybrid prosthetic replacement of knee joint using cement
O182	Conversion to hybrid prosthetic replacement of knee joint using cement
O183	Revision of hybrid prosthetic replacement of knee joint using cement
O184	Attention to hybrid prosthetic replacement of knee joint using cement
W400	Conversion from previous cemented total prosthetic replacement of knee joint
W402	Conversion to total prosthetic replacement of knee joint using cement
W403	Revision of total prosthetic replacement of knee joint using cement
W404	Revision of one component of total prosthetic replacement of knee joint using cement
W410	Conversion from previous uncemented total prosthetic replacement of knee joint
W412	Conversion to total prosthetic replacement of knee joint not using cement
W413	Revision of total prosthetic replacement of knee joint not using cement
W414	Revision of one component of total prosthetic replacement of knee joint not using cement
W420	Conversion from previous total prosthetic replacement of knee joint NEC
W422	Conversion to total prosthetic replacement of knee joint NEC

W423	Revision of total prosthetic replacement of knee joint NEC
W424*	Attention to total prosthetic replacement of knee joint NEC
W425	Revision of one component of total prosthetic replacement of knee joint NEC
W522†	Conversion to prosthetic replacement of articulation of bone using cement NEC
W523†	Revision of prosthetic replacement of articulation of bone using cement NEC
W532†	Conversion to prosthetic replacement of articulation of bone not using cement NEC
W533†	Revision of prosthetic replacement of articulation of bone not using cement NEC
W542†	Conversion to prosthetic replacement of articulation of bone NEC
W543†	Revision of prosthetic replacement of articulation of bone NEC
W544*†	Attention to prosthetic replacement of articulation of bone NEC
W553†	Conversion to prosthetic interposition arthroplasty of joint
W564†	Conversion to interposition arthroplasty of joint NEC
W574†	Conversion to excision arthroplasty of joint
W582†	Revision of resurfacing arthroplasty of joint
W603†	Conversion to arthrodesis and extra-articular bone graft NEC
W613†	Conversion to arthrodesis and articular bone graft NEC
W641†	Conversion to arthrodesis and internal fixation NEC
W642†	Conversion to arthrodesis and external fixation NEC
OPCS-4 co	odes for laterality

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

3
4
5
6
7
0
ð
9
10
11
12
13
14
15
16
17
10
18
19
20
21
22
23
24
25
26
20
27
28
29
30
31
32
33
34
35
36
20
3/
38
39
40
41
42
43
44
45
75 76
40
4/
48
49
50
51
52
53
54
55
22
50
57
58
59
60

1 2

Z942	Left-sided
Z943	Right-sided
ICD-10 co	odes for Infection
T845	Infection and inflammatory reaction due to internal joint prosthesis
T846	Infection and inflammatory reaction due to internal fixation device [any site]
T847	Infection and inflammatory reaction due to other internal orthopaedic prosthetic devices, implants and grafts
T814	Infection following a procedure, not elsewhere classified
ICD-10 co	odes for fracture
M966	Fracture of bone following insertion of orthopaedic implant, joint
	prosthesis or bone plate
ICD-10 co	odes for mechanical complications
T840	Mechanical complication of internal joint prosthesis
T841	Mechanical complication of internal fixation device of bones of limb
T842	Mechanical complication of internal fixation device of other bones
T843	Mechanical complication of other bone devices, implants and grafts
T844	Mechanical complication of other internal orthopaedic devices,
	imnplants and grafts
ICD-10 co	odes for osteoarthritis/arthrosis
M15-	Polyarthrosis
M17-	Gonarthrosis
M19-	Other arthrosis
OPCS-4 =	Office of Populations Censuses and Surveys Classification of
Interventio	ns and Procedures version 4. ICD-10 = International Statistical

Classification of Diseases and Related Health Problems, tenth revision. * Where

OPCS-4 codes Y032 (renewal of prosthesis in organ NOC) or Y037 (removal of prosthesis from organ NOC) were also used. † Where OPCS-4 codes O132 (knee NEC) or Z765 (lower end of femur NEC) or Z774 (upper end of tibia NEC) or Z787 (patella) or Z844 (patellofemoral joint) or Z845 (tibiofemoral joint) or Z846 (knee joint) or Z851 (upper tibiofibular joint) were used to identify knee as the body site.

for perteries only





77x73mm (300 x 300 DPI)

BMJ Open

Supplementary material S4 – R Code	
#Travel Times and Perioperative Outcomes in Revision Knee Replacement	
setwd("/Users/alexandermatthews//OneDrive - University of Exeter/Alex Mat MD/Revision Knee Networks MD/Travel Times Analysis_/")	thews
####Preparation of Data#### #load HES data	
RTKA2023 <- read.csv("~/Desktop/RTKA 06-09-23 CSV.csv")	
RTKA2023 <- read.csv("/Users/alexandermatthews//OneDrive - University of E Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/RTKA 06-0 CSV.csv")	xeter/Alex 19-23
#table only shows first 50 columns but we know there are 51 columns. Write t code to change preferences	his generic
rstudioapi::writeRStudioPreference("data_viewer_max_columns", 1000L)	
#Some entried are blank but are read as real values and not missing data #The table between age and sex shows three variables here #The dataset contains non standard missing values that are not recognised as #Replace empty strings with NA	NA
RTKA2023[RTKA2023 == ""] <- NA	
#Find number of incomplete cases in the data	
missing_data <- colSums(is.na(RTKA2023)) print(missing_data)	
#There are 14 entries with missing data only in the age group	
#check how many incomplete entries in age of patient column	
<pre>sum(!complete.cases(RTKA2023\$age_of_patient))</pre>	
#In case of missing values there are only 14 for age of patient #Can use imputation based on mean age #What is the mean age of the patients	

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

3
1
4
5
6
7
8
a
10
10
11
12
13
14
15
16
17
17
18
19
20
21
22
22
23
∠4 25
25
26
27
28
29
30
21
31
32
33
34
35
36
37
20
38
39
40
41
42
43
44
 1 E
45
46
47
48
49
50
51
51
52
53
54
55
56
57
52
50
59
60

1 2

mean(RTKA2023\$age_of_patient, na.rm = TRUE)

#mean age excluding missing values is 70
summary(RTKA2023\$age_of_patient, na.rm = TRUE)

#Check age is normally distributed

hist(RTKA2023\$age_of_patient)

#Input mean for missing values for age

RTKA2023\$age_of_patient[is.na(RTKA2023\$age_of_patient)] <- 69.82

#Now check number of missing values

sum(!complete.cases(RTKA2023\$age_of_patient))
#Now states 0 missing values

#There are other missing values for IMD decile ##In fact there are 439 IMD score missing values

sum(!complete.cases(RTKA2023\$IMD_score))

hist(RTKA2023\$IMD_score) #IMD score is non normally distributed

summary(RTKA2023\$IMD_score, na.rm = TURE)

#Median IMD score is 15.543

#Use imputation to impute median for missing value

RTKA2023\$IMD_score[is.na(RTKA2023\$IMD_score)] <- 15.543

#Check imputation complete

sum(!complete.cases(RTKA2023\$IMD_score))

#Now showing 0 missing values

#Next attach IMD decile number 6 to the missing values. As a score of 15 equates to the 6th decile

RTKA2023\$IMD_decile[is.na(RTKA2023\$IMD_decile)] <- 6

3	
4	
т 5	
5	
0	
/	
8	
9	
10	
11	
12	
13	
14	
15	
16	
10	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
20	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
20	
20	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
10	
47 50	
50	
51	
52	
53	
54	
55	
56	
57	
58	
50	
22	
οU	

#Check duplicate entry spells

duplicates <- RTKA2023[duplicated(RTKA2023),]

#No duplicates in data

#Frequencies of revisions by volume

as.numeric(RTKA2023\$TV12mo)

#frequencies of revisions by trust volume table(RTKA2023\$TVcat)

#Proportions by trust volume

prop.table(table(RTKA2023\$TVcat))

#Some entried are blank but are read as real values and not missing data #The table between age and sex shows three variables here #The dataset contains non standard missing values that are not recognised as NA #Replace empty strings with NA Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

RTKA2023[RTKA2023 == ""] <- NA

#Check this has registered

missing_data <- colSums(is.na(RTKA2023))
print(missing data)</pre>

#Column with LSOA_2011_Code has 171 missing.

#LSOA is part of primary exposure variable, small number of missing cases. Decision to remove rows rather than estimate from imputation because factor variable and dependent on provider code. Multiple imputation was used later to estimate missing travel data for these multiple rows where LSOA and site code was available

#Remove missing data in dataframe combined_data for column LSOA_2011_Code with missing fields = 171

RTKA2023<- RTKA2023[!is.na(RTKA2023\$LSOA_2011_Code),]

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

#16,565 patients before link with TRACC travel data

#Load Travel times data

TRAVELTIMES <- read.csv("~/Desktop/Drive time and Miles reference file.csv")

LSOAREF <- read.csv("~/Desktop/LSOA Matrix.csv")

LSOAREF <- read.csv("/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/LSOA Matrix.csv")

#Join data but The data is too big so we need to do this using SQL

install.packages("RSQLite") library(RSQLite)

query <- " Select * FROM times JOIN Isoa ON times.LSOAName = Isoa.LSOA11NM"

result <- dbGetQuery(con, query)

#10million 457 thousand and 999 possible combinations

#Write Dataframes

write.csv(result, "~/Desktop/JOINLSOATRAVEL.csv")

result<- read.csv("/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/JOINLSOATRAVEL.csv")

#####Now join this data to your revisions spreadsheet using key identifiers LSOA and Organisation site code

(query <- "
ç	Select *
F	FROM revisions3
J	IOIN travel3 ON revisions3.LSOA_2011_Code = travel3.LSOA11CD AND revisions3.Sitecode =
t	ravel3.ProviderSiteCode"
r	result_join <- dbGetQuery(con, query)
ŧ	#Number of patients following join 12,774
r	result1 <- result_join
Ŧ	Check your data for missing values
r	missing_data <- colSums(is.na(result1))
F	print(missing_data)
‡	#Check data for duplicates
(duplicates <- RTKA2023[duplicated(RTKA2023\$Epikey),]
‡	# Check for duplicates in the 'epikey' column
(duplicates <- result1[duplicated(result1\$Epikey),]
‡	#There are 2,047 duplicates
‡	#Remove duplicates in result 1
‡	# Remove duplicates: Keep only the first occurrence of each 'Epikey'
r	result1 <- result1[!duplicated(result1\$Epikey),]
‡	#final dataframe is 10,727
١	write.csv(result1, "/Users/alexandermatthews//OneDrive - University of Exeter/Alex
ſ	Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/FinalJOIN.csv")
‡	####Prepare Outcomes, Exposure variable and co-variates ####
,	*Cot up outcomos
1	rset up outcomes
1	

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

```
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

> result1\$Read30 <- ifelse(is.na(result1\$Read30), 'N', result1\$Read30) result1\$Read90 <- ifelse(is.na(result1\$Read90), 'N', result1\$Read90)

```
result1$Read30days <- ifelse(result1$Read30 == "Y", 1, 0)
#readmission for 90 days
result1$Read90days <- ifelse(result1$Read90 == "Y", 1, 0)</pre>
```

#Set up your co-variates

```
result1$HFRS_Band = as.factor(result1$HFRS_Band)
result1$HFRS_Band = relevel(result1$HFRS_Band, ref = 'None')
```

result1\$POD = as.factor(result1\$POD)
result1\$POD = relevel(result1\$POD, ref = 'EL')

table(result1\$POD)

#I've joined two dataframes based on a shared field. But some rows have not jointed

#Journey times statistics - 10,457,999 rows

#12,774 following join with revisions and travel data called "result1" but had duplicates 2,047 so remove these (duplicates due to slightly different latitude and longitude for same Site codes in journey times statistics)

#Final results 1 following removal of duplicates is 10,727

#Original dataframe is 16,736 called RTKA2023 following removal of early revisions, excluding missing LSOA was 16565

#Missing data for travel seen in 5,838 patients or 35% of patients

#Use multiple imputation to impute missing distance values for cases without join

#How many unmatched rows?

unmatched_rows <- RTKA2023[!(RTKA2023\$Epikey %in% result1\$Epikey),]

#There are 5,838 unmatched rows

#I want to create a dataframe showing both matched and unmatched fields based on this.

Identify columns that are in result1 but not in RTKA2023
missing_cols <- setdiff(names(result1), names(RTKA2023))</pre>

Add missing columns to RTKA2023 with NA values for (col in missing_cols) { RTKA2023[[col]] <- NA }
Ensure column order is the same as result1 RTKA2023 <- RTKA2023[, names(result1)]
Identify unmatched rows unmatched_rows <- RTKA2023[!(RTKA2023\$Epikey %in% result1\$Epikey),]
Combine matched rows (result1) with unmatched rows combined_data <- rbind(result1, unmatched_rows)
duplicates <- combined_data[duplicated(combined_data\$Epikey),]
#0 duplicates
write.csv(combined_data, "/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/FinalJOINCombined.csv")
combined_data <- read.csv("/Users/alexandermatthews//OneDrive - University of
Analysis_/FinalJOINCombined.csv")
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30)
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30) combined_data\$Read30days <- ifelse(combined_data\$Read30 == "Y", 1, 0)
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30) combined_data\$Read30days <- ifelse(combined_data\$Read30 == "Y", 1, 0) #Now have dataframe displaying both matched and unmatched rows
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30 combined_data\$Read30days <- ifelse(combined_data\$Read30 == "Y", 1, 0) #Now have dataframe displaying both matched and unmatched rows missing_data <- colSums(is.na(combined_data)) print(missing_data)
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30) combined_data\$Read30days <- ifelse(combined_data\$Read30 == "Y", 1, 0) #Now have dataframe displaying both matched and unmatched rows missing_data <- colSums(is.na(combined_data)) print(missing_data) #How many patients in high volume centres >49
Analysis_/FinalJOINCombined.csv") #Replace NA's in the Read columns with N combined_data\$Read30 <- ifelse(is.na(combined_data\$Read30), 'N', combined_data\$Read30) combined_data\$Read30days <- ifelse(combined_data\$Read30 == "Y", 1, 0) #Now have dataframe displaying both matched and unmatched rows missing_data <- colSums(is.na(combined_data)) print(missing_data) #How many patients in high volume centres >49 combined_data\$MRC <- ifelse(combined_data\$TV12mo > 49, 1, 0)

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

2	
3	
4	
5	
6	
0	
7	
8	
9	
10	
10	
11	
12	
13	
14	
14	
15	
16	
17	
10	
18	
19	
20	
21	
22	
22	
23	
24	
25	
25	
26	
27	
28	
20	
29	
30	
31	
32	
22	
22	
34	
35	
36	
27	
57	
38	
39	
40	
/1	
41	
42	
43	
44	
15	
45	
46	
47	
48	
⊿0	
77	
50	
51	
52	
52	
22	
54	
55	
56	
55	
5/	
58	
59	
<u> </u>	

1

nopatients <- subset(combined_data, MRC == 1)</pre> #6880 patients missing_data <- colSums(is.na(nopatients))</pre> print(missing data) # Count unique levels of ProvCode n_levels <- length(unique(nopatients\$ProvCode))</pre> cat("Number of unique providers (ProvCode):", n_levels, "\n") #38 providers #How many sites # Count unique levels of ProvCode n levels <- length(unique(nopatients\$Sitecode))</pre> cat("Number of unique sites (Sitecode):", n_levels, "\n") #187 sites #rates of readmission 30 days table(nopatients\$Read30days) #568/6880 8.3% #rates of mortality at 90 days table(nopatients\$Mort90days) #217/6880 3.2% #Rates of length of stay above median. Remember median calculated across entire cohort summary(combined data\$Spell Los) #Median of 5 nopatients\$Long_Los <- ifelse(nopatients\$Spell_Los > 5, 1, 0) table(nopatients\$Long_Los) #3421/6880 49.7% #3157 travel data not available #16,565 observations in entire dataframe not limited to teriatry referral centres #CV12mo missing 71 cases. Imputation using median due to positive skew

2	
3	
5	hist(combined_data\$CV12mo)
б	
7	#mean age excluding missing values is 70
8	summary(combined_data\$CV12mo, na.rm = TRUE)
9	
10	
12	#Input median of 6 for missing data
13	
14	combined data\$CV12mo[is.na(combined data\$CV12mo)] <- 6
15	
10	#Now need to use multiple imputation method to estimate travel data for columns
18	"DistanceMiles". "OffPeakDriveDistanceMiles". "PeakDriveTimes' based on associated
19	predictors:
20	
21	#Refer to this resource "https://bookdown.org/mwheymans/bookmi/multiple-
22	imputation html#setting_the_imputation_methods"
25 24	imputation.ntmi#setting-the-imputation-methods
25	#And this resource for contact
26	#And this resource for context
27	nttps://dept.stat.isa.umicn.edu/~jernck/courses/stat/01/notes/mi.ntmi
28	
29	# https://www.ebpi.uzh.ch/dam/jcr:dcUcef1/-29c/-4e61-8d33-
31	e690561ab7ae/mi_intro20191001.pdf (Advice on multi level modelling and imputation)
32	
33	# Install packages if they are not already installed
34	install.packages(c("mice", "ggplot2", "naniar"))
35	
36	# Load the packages
38	library(mice)
39	library(ggplot2)
40	library(naniar)
41	
42	#assuming missing data is due to random chance, LSOA and SiteCode are related to the
43	exposure but also include all other variables linked to your analysis
45	#Subset dataframe called combined date with only with relevant columns; age of patient.
46	sex. HFRS Band IMD Score. IMD Decile. infection. TVcat. CVcat. SiteCode. ProvCode. FinY.
47	DistanceMiles OffPeakDriveDistanceMiles PeakDriveTime Mort90days Read30 Spell Los
48	#decision not to include site code and LSOA as likely not present in missing data
49	"ISOA 2011 Code" "Sitecode"
50	
52	
53	
54	# Choosify the relevant columns live included TV/10 me as more by related to extend to
55	# specify the relevant columns i ve included 1V12mo as may be related to outcome,
56	Provide for clustering,
5/ 58	relevant_columns <- c(
50 59	"age_of_patient", "sex", "HFRS_Band", "IMD_score",
60	"infection", "TV12mo", "CV12mo", "ProvCode", "FinY",

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

2
3
4
-
5
6
7
8
0
9
10
11
12
12
15
14
15
16
17
17
18
19
20
21
21
22
23
24
25
25
26
27
28
20
29
30
31
32
22
22
34
35
36
27
57
38
39
40
⊿1
41
42
43
44
45
45
46
47
48
10
+7
50
51
52
52
55
54
55
56
57
57
58
59
60

1

"DistanceMiles", "OffPeakDriveDistanceMiles", "PeakDriveTime", "Mort90days", "Read30days", "Spell_Los"

)

Subset the dataframe with only the relevant columns subset_combined_data <- combined_data[, relevant_columns]</pre>

#Currently sex, HFRS_Band, TVCat, Sitecode, ProvCode, FinY are not incorporated in model as character variables

#convert these to factors

Convert variables to factors
subset_combined_data\$sex <- as.factor(subset_combined_data\$sex)
subset_combined_data\$ProvCode <- as.factor(subset_combined_data\$ProvCode)
subset_combined_data\$FinY <- as.factor(subset_combined_data\$FinY)
subset_combined_data\$HFRS_Band <- as.factor(subset_combined_data\$HFRS_Band)</pre>

subset_combined_data\$Sitecode <- as.factor(subset_combined_data\$Sitecode)
subset_combined_data\$LSOA_2011_Code <as.factor(subset_combined_data\$LSOA_2011_Code)</pre>

Check the structure of the dataframe to confirm str(subset_combined_data[, c("sex", "Sitecode", "ProvCode", "FinY", "HFRS_Band", "LSOA_2011_Code")])

#visualise missing data

vis_miss(subset_combined_data)

#35% missing travel data

Set the seed for reproducibility set.seed(123)

Perform Multiple Imputation

imp <- mice(subset_combined_data, m=5, method='pmm')</pre>

#Check for imputation values

imp\$imp\$OffPeakDriveDistanceMiles

#visualise imputed values
imp\$imp
#Means of the imputed values
imp\$chainMean
#What are the predictors
imp\$predictorMatrix
#Plot imputation values against observed values.
my_plot <- stripplot(imp, col=c("grey", "blue"), pch = c(1, 20))
my_plot
#Guidelines for imputation model suggest all variables in the analysis should be included, inclusive of dependent or outcome variables
#Ensure TVCat is not a predictor variable
pred <-imp\$predictorMatrix pred["TVcat"] <- 0 pred
#Plot the convergence (how equal is the variance to the mean)
plot(imp)
#Stack the imputed values into a single dataset and include original data
<pre>imp2 <- complete(imp, "long", inc = TRUE)</pre>
#Save imp2
write.csv(imp2, "/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/imp2.csv")
#Read it back in here:
imp2 <- read.csv("/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/imp2.csv")

BMJ Open

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

#Save as Supplemenatry figure

#Filter data by tertiary hospitals only

#But current guidelines suggest >49 is a high volume centre called a major revision centre and probably represents a unit with tertiary specialisation imp2\$MRC <- ifelse(imp2\$TV12mo > 49, 1, 0) tertiary revisions <- subset(imp2, MRC == 1)</pre> tertiary revisions\$Long Los <- ifelse(tertiary revisions\$Spell Los > 5, 1, 0) #declare the imputed data to be mids again, the format MICE is expecting for regression analyses tertiary revisions <- as.mids(tertiary revisions) #Now run your regression model using a multivariable model #A priori co-variates chosen based on evidence of predictors for readmission ####Primary Outcome 30 day readmission #### #Exposure 1 - Distance Miles library("lme4") # Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for clustering m3.mi <- with(tertiary revisions, glm(Read30days ~ DistanceMiles + IMD score + HFRS Band + sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode, family = "binomial")) print(m3.mi) # Pool results across imputed datasets pooled results <- pool(m3.mi) # Summarize pooled results with confidence intervals

Summarize pooled results with confidence intervals
summary_pooled <- summary(pooled_results, conf.int = TRUE)</pre>

Add Odds Ratios to the summary summary_pooled\$OR <- exp(summary_pooled\$estimate) summary_pooled\$Lower_CI <- exp(summary_pooled\$`2.5 %`) summary_pooled\$Upper_CI <- exp(summary_pooled\$`97.5 %`)
Display the final table with Odds Ratios and Confidence Intervals print(summary_pooled)
#check for evidence of multicollinearity?
library(car)
Use the long data including all imputations for VIF
<pre>tertiary_revisions <- complete(tertiary_revisions, "long", inc = TRUE)</pre>
<pre># Fit a logistic regression model on the complete dataset vif_model <- glm(Read30days ~ DistanceMiles + IMD_score + HFRS_Band + sex + age_of_patient + infection + TV12mo + CV12mo + FinY + ProvCode, data = tertiary_revisions, family = "binomial")</pre>
<pre># Calculate VIF vif_values <- vif(vif_model) print(vif_values)</pre>
#No evidence of multi-collinearity
#Is there a non linear relationship?
#Box Tidwell
#Recode back into correct format
tertiary_revisions <- as.mids(tertiary_revisions)
<pre># Custom function to add log-transformed variable and interaction term add_interaction <- function(data) { data\$Log_DistanceMiles <- log(data\$DistanceMiles) # Add log-transformed variable data\$Interaction <- data\$DistanceMiles * data\$Log_DistanceMiles # Add interaction term return(data) }</pre>

# Extract the long-format data including the original data	
tertiary_revisions_modified <- complete(tertiary_revisions, action = "long", include = TRL	JE)
# Apply the transformation to each imputed dataset	
tertiary_revisions_modified <- do.call("rbind",	
lapply(split(tertiary_revisions_modified,	
tertiary_revisions_modified\$.imp),	
# Convert back to mids object	
tertiary_revisions_modified <- as.mids(tertiary_revisions_modified)	
# Fit the logistic regression model with the interaction term	
model <- with(tertiary_revisions_modified, glm(Read30days ~ DistanceMiles + Interaction	n,
data = tert	
ramity = binomial(link = logit)))	
# Pool the results	
pooled_results <- pool(model)	
# Summarize pooled results	
summary_pooled <- summary(pooled_results, conf.int = TRUE)	
# Extract the n-value for the interaction term	
box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]	
# Print the p-value	
print(box_tidwen_p)	
# p value = 0.03 evidence of non linearity	
#Are spline terms significant for DistanceMiles if using 3 knots, 4 knots and 5 knots	
#Use data of all imputations in long format	
tertiary revisions <- complete(tertiary revisions, "long", inc = TRUE)	
# Load the required library	
library(splines)	
#AIC of non spline model	
model <- glm(Read30days ~ DistanceMiles, data = tertiary_revisions, family = binomial)	
summary(model)	

3	
4	HALC 21962
5	#AIC 21802
6	
7	# Define a function to fit and evaluate spline models with knots based on centiles
8	evaluate centile splines <- function(centiles data) {
9	# Calculate knots based on the specified contiles
10	
11	knots <- quantile(data\$DistanceMiles, probs = centiles, na.rm = TRUE)
12	
13	# Fit a logistic regression model with natural splines using the calculated knots
14	model_spline <- glm(Read30days ~ ns(DistanceMiles_knots = knots)
15	family = hinomial/link = "logit")
16	iamiy = binomiai(iink = logit),
17	data = data)
18	
19	# Summarize the model
20	summary model <- summary(model_spline)
21	summary_model < summary(model_spine)
22	
23	# Extract p-values for the spline terms
24	p_values <- summary_model\$coefficients[-1, "Pr(> z)"] # Exclude the intercept
25	
26	# Print the results
27	cat("\nPosults for contilos" contilos ".\n")
28	cat (investity for centiles , centiles,
29	print(p_values)
30	
3 I 2 2	# Return the model and calculated knots for further inspection if needed
2∠ 22	return(list(model = model_spline, p_values = p_values, knots = knots))
34	}
35	
36	# Evenue contile configurations for 2.4 and 5 limits
37	# Example centile configurations for 3, 4, and 5 knots
38	centiles_3_knots <- c(0.05, 0.50, 0.95) # 5th, 50th, and 95th percentiles
39	centiles_4_knots <- c(0.05, 0.35, 0.65, 0.95) # Custom centiles for 4 knots
40	centiles 5 knots <- c(0.05, 0.25, 0.50, 0.75, 0.95) # 5 knots centiles
41	(, , , , , , ,
42	# Evaluate medals with contile based knots using your detect
43	# Evaluate models with centile-based knots using your dataset
44	results_3_knots <- evaluate_centile_splines(centiles = centiles_3_knots, data =
45	tertiary_revisions)
46	results_4_knots <- evaluate_centile_splines(centiles = centiles_4_knots, data =
47	tertiary revisions)
48	results 5 knots $<$ - evaluate centile splines(centiles - centiles 5 knots data -
49	testing revisions)
50	
51	
52	# Compare models with centile-based knots
53	cat("\nComparing models with different centile-based knots:\n")
54	anova(results 3 knots\$model, results 4 knots\$model, results 5 knots\$model test =
55	"Chica")
50	
5/ 50	
50	# Print the calculated knot locations for each model
60	cat("\nKnot locations for 3 knots:\n")
~~	

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

print(results 3 knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results 4 knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results 5 knots\$knots) #AIC better fit 21806 #Model with 3 knots, significant terms but greater knots do not improve the model fit. Non linear relationship is evident and should be modelled with splines #Prepare predictors for model prediction #you need to ensure that the predicted probabilities align with the corresponding observations #Explore the data for missing values sum(!complete.cases(tertiary revisions\$DistanceMiles)) #Unimputed dataset is missing, so exclude these tertiary_revisions <- tertiary_revisions[!is.na(tertiary_revisions\$DistanceMiles),] sum(!complete.cases(tertiary_revisions\$sex)) sum(!complete.cases(tertiary revisions\$Read30days)) sum(!complete.cases(tertiary_revisions\$HFRS_Band)) sum(!complete.cases(tertiary revisions\$IMD score)) sum(!complete.cases(tertiary revisions\$infection)) #Currently infection as numeric - ensure is factor tertiary revisions\$infection <- as.factor(tertiary revisions\$infection) tertiary revisions\$HFRS Band <- as.factor(tertiary revisions\$HFRS Band) tertiary revisions\$sex <- as.factor(tertiary revisions\$sex)</pre> tertiary_revisions\$FinY <- as.factor(tertiary_revisions\$FinY)</pre> tertiary revisions\$ProvCode <- as.factor(tertiary revisions\$ProvCode) tertiary_revisions\$DistanceMiles <- as.numeric(tertiary_revisions\$DistanceMiles) tertiary revisions\$age of patient <- as.numeric(tertiary revisions\$age of patient) tertiary revisions\$IMD score <- as.numeric(tertiary revisions\$IMD score)

tertiary revisions\$TV12mo <- as.numeric(tertiary revisions\$TV12mo)

BMJ Open

2	
3	tertiary_revisions\$CV12mo <- as numeric(tertiary_revisions\$CV12mo)
4	
5	
6	#Run spline model with adjusted data excluding missing data
7	library(splines)
8	# For example, let's say you want 3 knots at specific percentiles
9	knots <- quantile(tertiary revisions\$DistanceMiles probs = $c(0.05, 0.50, 0.95)$ na rm =
10	$T_{\text{D}}(\mathbf{r}) = c(0.05, 0.50, 0.50), the set of t$
11	
12	print(knots)
13	#Knots at 53, 69 and 84
14	spline terms <- ns(tertiary revisions\$DistanceMiles, knots = knots)
15	
16	
17	
18	
19	model_with_custom_splines <- glm(Read30days ~ ns(DistanceMiles, knots = knots) +
20	HFRS Band + IMD score +
21	
22	family = "binamial" data = tartiary revisions)
23	ranniy – binonnar, uata – tertiary_revisions)
24	
25	
20	summary(model with custom splines)
27	
20	#Concrate a sequence of mean unit values for predicting
29	#Generate a sequence of mean unit values for predicting
30	
37	DistanceMiles_range <- seq(min(tertiary_revisions\$DistanceMiles),
32	max(tertiary revisions\$DistanceMiles), length.out = 100)
34	
35	new data <- expand grid(
36	Distance Miles - Distance Miles range
37	Distanceivilles = Distanceivilles_range,
38	sex = levels(tertiary_revisions\$sex), # Ensure it takes all factor levels
39	age_of_patient = mean(tertiary_revisions\$age_of_patient, na.rm = TRUE),
40	HFRS Band = levels(tertiary revisions\$HFRS Band), # Ensuring correct factor levels
41	IMD score = mean(tertiary revisionsSIMD score narm = TRUE)
42	FinV = lovels(tertiany_revisions\$1115_seere, name = neer,
43	rint – levels(lertiary_revisions; rint), # Ensuring correct factor levels
44	CV12mo = mean(tertiary_revisions\$CV12mo, na.rm = TRUE),
45	TV12mo = mean(tertiary_revisions\$TV12mo, na.rm = TRUE),
46	ProvCode = levels(tertiary_revisions\$ProvCode), # Ensuring correct factor levels
47	infection = levels(tertiary_revisionsSinfection) # Ensuring correct factor levels
48	
49	
50	
51	# Create a new dataset with a range of distances and miles and all other predictor variables
52	new_data <- expand.grid(DistanceMiles = DistanceMiles_range,
53	sex = unique(tertiary revisions\$sex),
54	age of patient = mean(tertiary revisions sage of natient)
55	HERC Rand - uniquo/tortiary_revisions¢HERC Rand)
50	IIIRS_Dallu – ullique(tertiary_revisiolispiners_Dallu),
5/	IND_score = mean(tertiary_revisions\$IMD_score),
20 20	FinY = unique(tertiary_revisions\$FinY),
59 60	CV12mo = mean(tertiary revisions\$CV12mo),
00	· · · · · · · · · · · · · · · · · · ·

Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

TV12mo = mean(tertiary_revisions\$TV12mo), infection = unique(tertiary_revisions\$infection))
Align the levels of ProvCode in new_data to match the training data new_data\$ProvCode <- factor(new_data\$ProvCode, levels = levels(tertiary_revisions\$ProvCode))
<pre># Align the levels of all relevant categorical variables new_data\$HFRS_Band <- factor(new_data\$HFRS_Band, levels = levels(tertiary_revisions\$HFRS_Band)) new_data\$sex <- factor(new_data\$sex, levels = levels(tertiary_revisions\$sex)) new_data\$FinY <- factor(new_data\$FinY, levels = levels(tertiary_revisions\$FinY)) new_data\$infection <- factor(new_data\$infection, levels = levels(tertiary_revisions\$infection))</pre>
#Factors are consistent with model
levels(new_data\$HFRS_Band) levels(tertiary_revisions\$HFRS_Band)
levels(new_data\$sex) levels(tertiary_revisions\$sex)
levels(new_data\$FinY) levels(tertiary_revisions\$FinY)
levels(new_data\$ProvCode) levels(tertiary_revisions\$ProvCode)
levels(new_data\$infection) levels(tertiary_revisions\$infection)
<pre># Check levels of ProvCode in both datasets setdiff(levels(new_data\$ProvCode), levels(tertiary_revisions\$ProvCode)) # Levels in new_data but not in tertiary_revisions setdiff(levels(tertiary_revisions\$ProvCode), levels(new_data\$ProvCode)) # Levels in tertiary_revisions but not in new_data</pre>
new_data\$ProvCode <- droplevels(new_data\$ProvCode) # Check for missing values in factor variables sum(is.na(new_data\$ProvCode)) # Number of missing values in ProvCode
Ensure that ProvCode is a factor new_data\$ProvCode <- factor(new_data\$ProvCode, levels = levels(tertiary_revisions\$ProvCode))

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

# Now try th	he prediction again
predicted_p	probs <- predict(model_with_custom_splines, newdata = new_data, type =
"response"))
# Combine r	mean_unit_range and predicted_probs into a data frame
plot_data <·	- data.frame(DistanceMiles = DistanceMiles_range, predicted_prob =
predicted_p	probs)
#Calculate 9	95% confidence intervals
# Obtain pre	edicted values and standard errors for the new data
predictions	<- predict(model_with_custom_splines, newdata = new_data, type = "link",
se.fit = TRU	E)
# Calculate f	the confidence intervals for the log-odds scale (link scale)
# Use a 95%	6 confidence level (z-value = 1.96 for a 95% CI)
z_value <- 1	1.96
log_odds_lc	ower <- predictions\$fit - z_value * predictions\$se.fit
log_odds_u	pper <- predictions\$fit + z_value * predictions\$se.fit
# Convert th	ne log-odds confidence intervals to probabilities
# First, appl [,]	y the inverse link function (logistic function) to the log-odds
lower_prob	<- plogis(log_odds_lower)
upper_prob	o <- plogis(log_odds_upper)
<pre># Combine t plot_data <- DistanceM predicted_ ci_lower = ci_upper =)</pre>	the predicted probabilities and their confidence intervals into a data frame - data.frame(liles = new_data\$DistanceMiles, _prob = plogis(predictions\$fit), # Logistic transformation of the link lower_prob, upper_prob
# Combine r plot_data <-	mean_unit_range, predicted_probs, ci_lower, and ci_upper into plot_data - data.frame(DistanceMiles = DistanceMiles_range, predicted_prob = predicted_probs, ci_lower = boot_results\$ci_lower, ci_upper = boot_results\$ci_upper)
library(ggplo	ot2)
# Plot the sp	pline curve with confidence intervals
ggplot(plot_	_data, aes(x = DistanceMiles)) +
```
2
3
              geom line(aes(y = predicted prob), color = "blue", size = 1) +
4
              geom ribbon(aes(ymin = ci lower, ymax = ci upper), fill = "blue", alpha = 0.2) +
5
              labs(x = "Distance (Miles)", y = "Predicted Probability Readmission at 30 days") +
6
7
              theme minimal()
8
9
             library(dplyr)
10
11
             # Group by mean unit and calculate mean predicted prob and corresponding confidence
12
13
             intervals
14
             mean data <- plot data %>%
15
              group by(DistanceMiles) %>%
16
              summarise(
17
18
               mean predicted prob = mean(predicted prob, na.rm = TRUE),
19
               mean ci lower = mean(ci lower, na.rm = TRUE),
20
               mean ci upper = mean(ci upper, na.rm = TRUE)
21
              )
22
23
24
25
             # Define specific breaks (e.g., 25, 50, 75, ..., up to the maximum)
26
             breaks seq <- seq(0, max(mean data$DistanceMiles, na.rm = TRUE), by = 5)
27
28
29
             library(ggplot2)
30
             # Plot with specified increments on x-axis
31
             ggplot(mean_data, aes(x = DistanceMiles, y = mean_predicted_prob)) +
32
              geom point() + # Add points for mean predicted prob
33
              geom line() + # Connect points with a line
34
35
              geom_ribbon(aes(ymin = mean_ci_lower, ymax = mean_ci_upper), fill = "blue", alpha =
36
             0.2) + # Add ribbon for confidence intervals
37
              labs(x = "Travel Distance (Miles)", y = "Mean Predicted Probability for readmission at 30
38
             days", title = "Spline curve predicted probability of readmission at 30 days by patient travel
39
             distance") +
40
41
              scale x continuous(limits = c(0, max(mean data$DistanceMiles, na.rm = TRUE)), breaks =
42
             breaks seq) +
43
              theme minimal() +
44
              theme(
45
46
               axis.title.x = element text(size = 14), # Increase x-axis title font size
47
               axis.title.y = element text(size = 14), # Increase y-axis title font size
48
               axis.text.x = element text(size = 12), # Increase x-axis tick label font size
49
               axis.text.y = element text(size = 12), # Increase y-axis tick label font size
50
               plot.title = element text(size = 16, hjust = 0.5) # Increase plot title font size and center it
51
52
              )
53
54
55
             #Spline curve does appear to show the predicted probability of emergency readmission at
56
57
             30 days increases with travel distance but wide confidence intervals
58
59
             #Model Distance Miles and 30 day readmission with 3 knot splines
60
```

####First Imputation and descriptive stats####
#Use first imputed data for clinical and demographic characteristic summary
#complete_data is the first imputation
Count unique levels of ProvCode n_levels <- length(unique(complete_data\$ProvCode)) cat("Number of unique providers (ProvCode):", n_levels, "\n")
Count unique levels of sites n_levels <- length(unique(complete_data)) cat("Number of unique providers (ProvCode):", n_levels, "\n")
Count unique levels of ProvCode n_levels <- length(unique(tertiary_revisions\$ProvCode)) cat("Number of unique providers (ProvCode):", n_levels, "\n")
#38 unique providers
#Number of sites
Count unique levels of Sites but need to use original dataframe as sites not included in imputation analysis
#Find all those attending tertirary referral centre from original data tertiary_all <- subset(combined_data, MRC == 1)
#Find number of sites n_levels <- length(unique(tertiary_all\$Sitecode)) cat("Number of unique providers (Sites):", n_levels, "\n")
#187 sites
#Back to first imputation dataset. Calculate median number of miles straight line distance
summary(complete_data\$DistanceMiles)
#Median is 7.1 IQR is 3.9 to 12.7. Range 0 to 77.1 miles.

distance

#Driving distances

#Median 10.4 miles, IQR is 5.8 to 18.3 miles #Calculate median driving times summary(complete_data\$PeakDriveTime) #Median is 27 minutes IQR is 18.4 to 38.4. Maximum 104 minutes #Create travel time quintile variable quintiles <- quantile(complete_data\$DistanceMiles, probs = seq(0,1,0.2), na.rm=TRUE) complete data\$distancequintile <- cut(complete data\$DistanceMiles, breaks = quintiles, ..lowe labels = c("Q1", "Q2", "Q3", "Q4", "Q5"), include.lowest = TRUE) #Tabulate descriptive stats hist(tertiary_all\$Spell_Los) summary(tertiary_all\$Spell_Los) # Total number of revisions total_revisions <- nrow(complete_data)</pre> # Create a summary table summary_stats <- complete_data %>% group by(distancequintile) %>% summarise(# Count of observations Count = n(), **#** Distinct Providers Distinct Units = n distinct(ProvCode), Total Distinct Units = n distinct(complete data\$ProvCode), Distinct_Units_Percent = (Distinct_Units / Total_Distinct_Units) * 100, #Median distance Distance LowerQuartile = quantile(DistanceMiles, 0.25, na.rm = TRUE), Distance Median = median(DistanceMiles, na.rm = TRUE), Distance UpperQuartile = quantile(DistanceMiles, 0.75, na.rm = TRUE),

summary(complete data\$OffPeakDriveDistanceMiles)

2	
3	#Maan driving time
4	
5	DrivingTime_LowerQuartile = quantile(PeakDriveTime, 0.25, na.rm = TRUE),
6	DrivingTime_Median = median(PeakDriveTime, na.rm = TRUE),
7	DdrivingTime UpperQuartile = quantile(PeakDriveTime, 0.75, na.rm = TRUE),
8	<u> </u>
9	
10	
11	
12	# Age: Mean and standard deviation
13	Age Mean = mean(age of patient, na.rm = TRUE).
14	Age SD = sd(age of patient na rm = TRLE)
15	$Agc_{5D} = 30(agc_{0} patient, na. in = inoc),$
16	
17	# Age: Mean ± SD (concatenated)
18	Age_Mean_SD = paste(round(mean(age_of_patient, na.rm = TRUE), 2), "±",
19	round(sd(age_of_patient, na.rm = TRUE), 2)).
20	
21	
22	
23	# Gender: frequency and percentage
24	Female Freq = sum(sex == "Female", na.rm = TRUE),
25	Female_Percent = sum(sex == "Female", na.rm = TRUE) / n() * 100.
26	Malo $Frog = sum(sov == "Malo" na rm = TPLIE)$
27	$Male_neq = sum(sex = - male , malin = mol(s),$
28	Male_Percent = sum(sex == "Male", na.rm = TRUE) / n() * 100,
29	
30	# ASA: frequency and percentage for each level
31	HFRS None Freg = sum(HFRS Band == "None", na.rm = TRUE).
32	HEPS Nono Dercent = sum (HEPS Band == "Nono" na rm = TPLIE) / n() * 100
33	HERC Mild Free a willer Pard In Mild a True True (
34	HFRS_IVIIId_Freq = sum(HFRS_Band == IVIIId , na.rm = IRUE),
35	HFRS_Mild_Percent = sum(HFRS_Band == "Mild", na.rm = TRUE) / n() * 100,
36	HFRS_Moderate_Freq = sum(HFRS_Band == "Moderate", na.rm = TRUE),
3/	HFRS Moderate Percent = sum(HFRS Band == "Moderate", na.rm = TRUE) / n() * 100.
38	HERS Severe Freq - sum/HERS Band "Severe" na rm - TRUE)
39	HERC Covere Devent over (UERC Deed - "Covere" of the TRUE) (r() * 100
40	HFRS_Severe_Percent = sum(HFRS_Band == Severe , na.rm = TRUE) / n() * 100,
41	
42	
45	#Infection
44	
45	Infaction From - cum/infaction "1" no rm - TDUE)
40	infection_Freq = sum(infection == 1, na.rm = IROE),
47	Infection_Percent = sum(infection == "1", na.rm = TRUE) / n() * 100,
40	
50	
51	# Year: frequency and percentage for each year from 2009 to 2019
52	Vor 2015 2016 Erog = $sum (EinV = - "201E / 16" no rm = TBUE)$
53	Teal_2015_2010_FIEY = SUIT(FILIT == 2015/10, Hd.HII = TRUE),
54	rear_2015_2016_Percent = sum(Finr == "2015/16", na.rm = TRUE) / n() * 100,
55	Year_2016_2017_Freq = sum(FinY == "2016/17", na.rm = TRUE),
56	Year_2016_2017_Percent = sum(FinY == "2016/17", na.rm = TRUE) / n() * 100,
57	Year 2017 2018 Freq = sum(FinY == "2017/18", na rm = TRUF).
58	Very 2017 2018 Dercent - $sum/EinV - "2017/19"$ no rm - TPLIEV / n/) * 100
59	$100_{201}/2010_{100}$ = 100_{100}
60	rear_2018_2019_Freq = sum(Finr == $2018/19^{\circ}$, na.rm = IRUE),

Year_2018_2019_Percent = sum(FinY == "2018/19", na.rm = TRUE) / n() * 100, Year_2019_2020_Freq = sum(FinY== "2019/20", na.rm = TRUE), Year_2019_2020_Percent = sum(FinY == "2019/20", na.rm = TRUE) / n() * 100,
Median Surgeon Volume: lower quartile, median, and upper quartile Surgeon_LowerQuartile = quantile(CV12mo, 0.25, na.rm = TRUE), Surgeon_Median = median(CV12mo, na.rm = TRUE), Surgeon_UpperQuartile = quantile(CV12mo, 0.75, na.rm = TRUE),
#Median hospital volume
Hospital_LowerQuartile = quantile(TV12mo, 0.25, na.rm = TRUE), Hospital_Median = median(TV12mo, na.rm = TRUE), Hospital_UpperQuartile = quantile(TV12mo, 0.75, na.rm = TRUE),
#Median IMD Score
IMD_LowerQuartile = quantile(IMD_score, 0.25, na.rm = TRUE), IMD_Median = median(IMD_score, na.rm = TRUE), IMD_UpperQuartile = quantile(IMD_score, 0.75, na.rm = TRUE),
Print the summary table print(summary_stats)
write.csv(summary_stats, "/Users/alexandermatthews//OneDrive - University of Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times Analysis_/Summary_stats.csv")
####Cluster Variable ####
<pre>####Cluster Variable #### # Compute the mean outcome for each cluster library(dplyr) prov_means <- tertiary_revisions %>% group_by(ProvCode) %>% summarize(mean_outcome = mean(Read30days, na.rm = TRUE)) # Plot variability</pre>
<pre>####Cluster Variable #### # Compute the mean outcome for each cluster library(dplyr) prov_means <- tertiary_revisions %>% group_by(ProvCode) %>% summarize(mean_outcome = mean(Read30days, na.rm = TRUE)) # Plot variability</pre>

2			
3	hoxplot(mean_outcome ~ ProvCode_data = prov_means_xlab = "ProvCode"_vlab = "Mean		
4			
5	Outcome")		
6			
7	# Summary statistics of variability		
8			
0	summary(prov_means\$mean_outcome)		
9 10			
10	#There is evidence of variability between providers		
11	in there is evidence of valiability between providers		
12			
13			
14	# Fit logistic regression on imputed datasets		
15	m2 mi < with/tentient, revisions, almer/Deed20deve & DistanceMiles + IMD, seens +		
16	m3.ml <- with(tertiary_revisions, gimer(kead30days ~ Distanceivilies + liviD_score +		
17	HFRS_Band +		
18	sex + age of patient + infection + TV12mo + CV12mo + FinY + (1		
19			
20	Proveodej,		
20	family = "binomial"))		
21			
22			
23			
24	print(m3.mi)		
25			
26	Hinduding DrayCode as a random affect was tasted but led to convergence issues likely due		
27	#including Proveoue as a random effect was tested but led to convergence issues likely due		
28	to numerical instability between providers so a decision was made to accept the fixed		
29	effects model which may account for clustering at the provider level but is a limitation of		
30	the study		
31			
37			
22			
22	#Was travel distance strengly correlated with IMD score or ago?		
34	#was travel distance strongly correlated with hvid_score of age:		
35			
36			
37			
38			
39	#Next do a Spearman's rank correlation between travel distance and age, and then for		
40	travel distance and IMD score		
41			
42			
43	imp2\$MRC <- ifelse(imp2\$1V12mo > 49, 1, 0)		
44			
15	tertiary revisions <- subset(imn2_MBC == 1)		
45	$certary_revisions < subset(mp2, where = 1)$		
40			
4/			
48	write csy/tertiary_revisions_"/Lisers/alexandermatthews//OneDrive - University of		
49	Fustor (Alex Matthews MD / Devision Knoc Naturalis MD / Travel Times		
50	Exeter/Alex Matthews MD/Revision Knee Networks MD/Travel Times		
51	Analysis_/tertiary_revisions.csv")		
52			
53			
54			
55	tertiary_revisions <- as.mids(tertiary_revisions)		
56			
57	tortions revisions for a patient a		
57	ter tiar y_revisionspage_or_patient <-		
00	as.numeric(as.character(tertiary_revisions\$age_of_patient))		
59			
60			

tertiary revisions\$DistanceMiles <as.numeric(as.character(tertiary revisions\$DistanceMiles)) #Age and travel distance, Cannot pool the results based on the multiple imputations as cor test not compatible. Therefore stack all imputations together and calculate correlation # Scatterplot with linear regression line plot(tertiary revisions\$age of patient, tertiary revisions\$DistanceMiles, main = "Scatterplot of Age of Patient vs DistanceMiles", xlab = "Age of Patient", ylab = "DistanceMiles", pch = 19, col = "blue"# Add a linear trendline abline(Im(DistanceMiles ~ age of patient, data = tertiary revisions), col = "red", lwd = 2) # Calculate Spearman's rank correlation spearman test <- cor.test(tertiary revisions\$age of patient, tertiary revisions\$DistanceMiles, method = "spearman") # Extract rho and p-value rho <- round(spearman_test\$estimate, 2)</pre> p value <- spearman test\$p.value p_value_text <- ifelse(p_value < 0.05, "<0.05", paste0("=", round(p_value, 3)))</pre> # Add a legend with Spearman's rank correlation information legend("topright", legend = paste("Spearman's Rank Correlation:\n", "rho =", rho, ", p-value", p_value_text), col = c("blue", "red"), lty = c(NA, 1), pch = c(19, NA), lwd = c(NA, 2), bty = "n") #IMD score and travel distance # Scatterplot with trendline plot(tertiary revisions\$IMD score, tertiary revisions\$DistanceMiles, main = "Scatterplot of IMD score vs DistanceMiles", xlab = "IMD score", ylab = "DistanceMiles", pch = 19, col = "blue"# Add a linear trendline (for visualizing the general trend) abline(Im(DistanceMiles ~ IMD score, data = tertiary revisions), col = "red", lwd = 2) # Calculate Spearman's rank correlation spearman_test <- cor.test(tertiary_revisions\$IMD_score, tertiary_revisions\$DistanceMiles, method = "spearman")

2			
3			
4	# Extract rho and p-value		
5	rho < round(snoorman_tost¢ostimato_2)		
6	The conductive formation in the second		
/	p_value <- spearman_test\$p.value		
8	p_value_text <- ifelse(p_value < 0.05, "<0.05", paste0("=", round(p_value, 3)))		
9			
10	# Add a legend with Spearman's rank correlation information		
12	legend("topright", legend = paste("Spearman's Rank Correlation:\n",		
13	"rho -" rho " n-value" n value text)		
14	rol = c("blue" "rod") [ty = c(NA 1) rob = c(10 NA) by d = c(NA 2) bty = "p")		
15	cor = c(brue, red), red = c(nA, 1), pcr = c(19, nA), rwd = c(nA, 2), bry = rr)		
16			
17	#Exposure 2 - OffPeakDriveDistanceMiles		
18			
19	library("Ime4")		
20			
21	# Eit logistic regression on imputed datasets include ProvCode in fixed effects to account for		
22	# The logistic regression on impliced datasets include Proveode in fixed effects to account for		
23	clustering		
24	m3.mi <- with(tertiary_revisions, glm(Read30days ~ OffPeakDriveDistanceMiles +		
25	IMD_score + HFRS_Band +		
20	sex + age_of_patient + infection + TV12mo + CV12mo + FinY +		
27	ProvCode.		
20	family = "hinomial"))		
30			
31			
32			
33	print(m3.mi)		
34			
35			
36	# Pool results across imputed datasets		
37	nooled results <- nool(m3 mi)		
38			
39			
40	# Summarize pooled results with confidence intervals		
41	summary_pooled <- summary(pooled_results, conf.int = TRUE)		
42 43			
44	# Add Odds Ratios to the summary		
45	summary pooledSOR <- exp(summary pooledSestimate)		
46	summary pooled $(-exp(summary pooled))$		
47	$summary_pooleustower_cl < exp(summary_pooleus 2.5 /0)$		
48	summary_pooleu\$opper_cr <- exp(summary_pooleu\$ 37.5 %)		
49			
50	# Display the final table with Odds Ratios and Confidence Intervals		
51	print(summary_pooled)		
52			
53	#check for evidence of multicollinearity?		
54	, ,		
55	library(car)		
50 57			
58 58			
50	# Use the first imputed dataset for the VIF calculation		
60	complete_data <- complete(tertiary_revisions, 1)		

```
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

```
# Fit a logistic regression model on the complete dataset
vif model <- glm(Read30days ~ OffPeakDriveDistanceMiles + IMD score + HFRS Band +
          sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode,
         data = complete data, family = "binomial")
# Calculate VIF
vif values <- vif(vif model)
print(vif_values)
#No evidence of multi-collinearity
#Is there a non linear relationship?
# Custom function to add log-transformed variable and interaction term
add interaction <- function(data) {
 data$Log OffPeakDriveDistanceMiles <- log(data$OffPeakDriveDistanceMiles) # Add log-
transformed variable
 data$Interaction <- data$OffPeakDriveDistanceMiles *</pre>
data$Log_OffPeakDriveDistanceMiles # Add interaction term
 return(data)
}
# Extract the long-format data including the original data
tertiary_revisions_modified <- complete(tertiary_revisions, action = "long", include = TRUE)
# Apply the transformation to each imputed dataset
tertiary revisions modified <- do.call("rbind",
                     lapply(split(tertiary_revisions_modified,
tertiary revisions modified$.imp),
                         add interaction))
# Convert back to mids object
tertiary_revisions_modified <- as.mids(tertiary_revisions_modified)
# Fit the logistic regression model with the interaction term
model <- with(tertiary revisions modified, glm(Read30days ~ OffPeakDriveDistanceMiles +
Interaction,
                          family = binomial(link = "logit")))
# Pool the results
pooled results <- pool(model)
# Summarize pooled results
summary_pooled <- summary(pooled_results, conf.int = TRUE)</pre>
```

# Extract the p-value for the inter box_tidwell_p <- summary_poole	action term d[summary_pooled\$term == "Interaction", "p.value"]
# Print the p-value print(box_tidwell_p)	
# p value = 0.05. There is no evide	ence of non linearity
#Exposure 3 - PeakDriveTime	
library("Ime4")	
# Fit logistic regression on impute clustering	ed datasets include ProvCode in fixed effects to account for
m3.mi <- with(tertiary_revisions, HFRS_Band +	glm(Read30days ~ PeakDriveTime + IMD_score +
sex + age_c ProvCode,	of_patient + infection + TV12mo + CV12mo + FinY +
, family = "bir	nomial"))
print(m3.mi)	
# Pool results across imputed dat pooled_results <- pool(m3.mi)	asets
# Summarize pooled results with summary_pooled <- summary(po	confidence intervals oled_results, conf.int = TRUE)
# Add Odds Ratios to the summar summary_pooled\$OR <- exp(sum summary_pooled\$Lower_CI <- ex summary_pooled\$Upper_CI <- e>	Y mary_pooled\$estimate) (p(summary_pooled\$`2.5 %`) (p(summary_pooled\$`97.5 %`)
# Display the final table with Odd print(summary_pooled)	s Ratios and Confidence Intervals
#check for evidence of multicollin	earity?
library(car)	
# Use the first imputed dataset fo complete_data <- complete(tertia	or the VIF calculation ary_revisions, 1)

```
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

```
# Fit a logistic regression model on the complete dataset
vif model <- glm(Read30days ~ PeakDriveTime + IMD score + HFRS Band +
          sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode,
         data = complete data, family = "binomial")
# Calculate VIF
vif values <- vif(vif model)
print(vif_values)
#No evidence of multi-collinearity
#Is there a non linear relationship?
# Custom function to add log-transformed variable and interaction term
add interaction <- function(data) {
 data$Log PeakDriveTime <- log(data$PeakDriveTime) # Add log-transformed variable
 data$Interaction <- data$PeakDriveTime * data$Log PeakDriveTime # Add interaction
term
 return(data)
}
# Extract the long-format data including the original data
tertiary_revisions_modified <- complete(tertiary_revisions, action = "long", include = TRUE)
# Apply the transformation to each imputed dataset
tertiary revisions modified <- do.call("rbind",
                     lapply(split(tertiary_revisions_modified,
tertiary revisions modified$.imp),
                         add interaction))
# Convert back to mids object
tertiary_revisions_modified <- as.mids(tertiary_revisions_modified)
# Fit the logistic regression model with the interaction term
model <- with(tertiary_revisions_modified, glm(Read30days ~ PeakDriveTime + Interaction,
                          family = binomial(link = "logit")))
# Pool the results
pooled results <- pool(model)
# Summarize pooled results
summary_pooled <- summary(pooled_results, conf.int = TRUE)</pre>
```

2		
3		
4	# Extract the p-value for the interaction term	
5	box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]	
6		
7	# Print the p-value	
8	print/box tidwell n)	
9	print(box_tidweil_p)	
10		
11	# p value = 0.13 not evidence of non linearity	
12		
13		
14		
15		
16		
17		
18	####Secondary Outcome mortality 90 days ####	
19		
20		
21	#Evnesure 1 Distance Miles	
22	#Exposure 1 - Distance Miles	
23		
24	library("lme4")	
25		
26	# Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for	
27	clustoring	
28		
29	m3.ml <- with(tertiary_revisions, glm(Mort90days ~ DistanceMiles + IMD_score +	
30	HFRS_Band +	
31	sex + age_of_patient + infection + TV12mo + CV12mo + FinY +	
2∠ 22	ProvCode.	
34	family = "hinomial"))	
35		
36		
37	La	
38	print(m3.mi)	
39		
40		
41	# Pool results across imputed datasets	
42	noolod rosults <- nool(m3 mi)	
43	pooled_results <- pool(ins.ini)	
44		
45	# Summarize pooled results with confidence intervals	
46	summary_pooled <- summary(pooled_results, conf.int = TRUE)	
47		
48	# Add Odds Ratios to the summary	
49	π rad odds hallos to the summary peopled (estimate)	
50	summary_pooled\$UK <- exp(summary_pooled\$estimate)	
51	summary_pooled\$Lower_Cl <- exp(summary_pooled\$ 2.5 %`)	
52	summary_pooled\$Upper_CI <- exp(summary_pooled\$`97.5 %`)	
53		
54	# Display the final table with Odds Ratios and Confidence Intervals	
55 56	nrint(summary_nooled)	
50 57		
57 50		
50	#check for evidence of multicollinearity?	
5 9		
00		

```
library(car)
# Use the first imputed dataset for the VIF calculation
complete_data <- complete(tertiary_revisions, 1)
# Fit a logistic regression model on the complete dataset
vif model <- glm(Mort90days ~ DistanceMiles + IMD score + HFRS Band +
          sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode,
         data = complete_data, family = "binomial")
# Calculate VIF
vif values <- vif(vif model)
print(vif_values)
#No evidence of multi-collinearity
#Is there evidence of non linearity?
library(mice)
tertiary_revisions <- as.mids(tertiary_revisions)
# Custom function to add log-transformed variable and interaction term
add interaction <- function(data) {
 data$Log DistanceMiles <- log(data$DistanceMiles) # Add log-transformed variable
 data$Interaction <- data$DistanceMiles * data$Log DistanceMiles # Add interaction term
 return(data)
}
# Extract the long-format data including the original data
tertiary revisions modified <- complete(tertiary revisions, action = "long", include = TRUE)
# Apply the transformation to each imputed dataset
tertiary revisions modified <- do.call("rbind",
                     lapply(split(tertiary_revisions_modified,
tertiary revisions modified$.imp),
                         add interaction))
# Convert back to mids object
tertiary revisions modified <- as.mids(tertiary revisions modified)
# Fit the logistic regression model with the interaction term
model <- with(tertiary revisions modified, glm(Mort90days ~ DistanceMiles + Interaction,
                          family = binomial(link = "logit")))
```

Pool the results pooled_results <- pool(model)
Summarize pooled results summary_pooled <- summary(pooled_results, conf.int = TRUE)
Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]
Print the p-value print(box_tidwell_p)
P value 0.95
#Exposure 2 - OffPeakDriveDistanceMiles
library("lme4")
Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for clustering m3.mi <- with(tertiary revisions, glm(Mort90days ~ OffPeakDriveDistanceMiles +
IMD_score + HFRS_Band + sex + age_of_patient + infection + TV12mo + CV12mo + FinY +
ProvCode, family = "binomial"))
print(m3.mi)
Pool results across imputed datasets pooled_results <- pool(m3.mi)
Summarize pooled results with confidence intervals summary_pooled <- summary(pooled_results, conf.int = TRUE)
Add Odds Ratios to the summary summary_pooled\$OR <- exp(summary_pooled\$estimate) summary_pooled\$Lower_CI <- exp(summary_pooled\$`2.5 %`) summary_pooled\$Upper_CI <- exp(summary_pooled\$`97.5 %`)
Display the final table with Odds Ratios and Confidence Intervals print(summary_pooled)

1 2	
3	
4	#check for evidence of multicollinearity?
5	
7	librarv(car)
8	
9	# Use the first imputed dataset for the VIE calculation
10	complete data <- complete(tertiary revisions 1)
11	
13	# Fit a logistic regression model on the complete dataset
14	vif model <- glm(Read30days ~ OffPeakDriveDistanceMiles + IMD_score + HERS_Band +
15	sex + age of natient + infection + TV12mo + CV12mo + FinY + ProvCode
16	data = complete_data_family = "hinomial")
17	
19	# Calculate VIE
20	vif values <- vif(vif model)
21	nrint(vif_values)
22	
23	
25	#No evidence of multi-collinearity
26	
27	#Is there evidence of non linearity?
28 29	institute evidence of non-incurry.
30	tertiary revisions <- as mids(tertiary revisions)
31	
32	# Custom function to add log-transformed variable and interaction term
33	add interaction <- function(data) {
35	data\$1.0g_OffPeakDriveDistanceMiles <= log(data\$OffPeakDriveDistanceMiles) # Add log_
36	transformed variable
37	dataSinteraction <- dataSOffPeakDriveDistanceMiles *
38	data\$l.og_OffPeakDriveDistanceMiles_#Add_interaction_term
39 40	return(data)
41	
42	
43	# Extract the long-format data including the original data
44 45	tertiary revisions modified <- complete/tertiary revisions action = "long" include = TRUE)
45 46	
47	# Apply the transformation to each imputed dataset
48	π Apply the transformation to each imputed dataset
49	lannly/snlit/tertiary revisions modified
50 51	tortiary rovisions modified\$ imp)
52	add interaction)
53	
54	# Convert back to mids object
55	π convert back to milds object tertiary revisions modified z_{-} as mids/tertiary revisions modified)
56 57	
58	# Fit the legistic regression model with the interaction term
59	
60	

model <- with(tertiary_revisions_modified, glm(Mort90days ~ OffPeakDriveDistanceMiles + Interaction.
family = binomial(link = "logit")))
Pool the results pooled_results <- pool(model)
Summarize pooled results summary_pooled <- summary(pooled_results, conf.int = TRUE)
Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]
Print the p-value print(box_tidwell_p)
#0.989
#Exposure 3 - PeakDriveTime
library("lme4")
Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for clustering
m3.mi <- with(tertiary_revisions, glm(Mort90days ~ PeakDriveTime + IMD_score + HFRS_Band +
sex + age of patient + infection + TV12mo + CV12mo + FinY +
ProvCode,
family = "binomial"))
print(m3.mi)
Pool results across imputed datasets pooled_results <- pool(m3.mi)
Summarize pooled results with confidence intervals summary_pooled <- summary(pooled_results, conf.int = TRUE)
Add Odds Ratios to the summary summary_pooled\$OR <- exp(summary_pooled\$estimate) summary_pooled\$Lower_CI <- exp(summary_pooled\$`2.5 %`)

```
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

summary_pooled\$Upper_Cl <- exp(summary_pooled\$`97.5 %`)</pre>

Display the final table with Odds Ratios and Confidence Intervals
print(summary_pooled)

#check for evidence of multicollinearity?

library(car)

Use the first imputed dataset for the VIF calculation complete_data <- complete(tertiary_revisions, 1)</pre>

Calculate VIF
vif_values <- vif(vif_model)
print(vif_values)</pre>

#No evidence of multi-collinearity

```
#Is there evidence of non linearity?
```

```
# Custom function to add log-transformed variable and interaction term
add_interaction <- function(data) {
    data$Log_PeakDriveTime <- log(data$PeakDriveTime) # Add log-transformed variable
    data$Interaction <- data$PeakDriveTime * data$Log_PeakDriveTime # Add interaction</pre>
```

term

return(data)

}

Extract the long-format data including the original data tertiary_revisions_modified <- complete(tertiary_revisions, action = "long", include = TRUE)</pre>

```
# Convert back to mids object
tertiary_revisions_modified <- as.mids(tertiary_revisions_modified)</pre>
```

Fit the logistic regression model with the interaction term

model <- with(tertiary_revisions_modified, glm(Mort90days ~ PeakDriveTime + Interaction, family = binomial(link = "logit")))
Pool the results pooled_results <- pool(model)
Summarize pooled results summary_pooled <- summary(pooled_results, conf.int = TRUE)
Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]
Print the p-value print(box_tidwell_p)
P avlue 0.78
####Secondary outcome prolonged LOS ####
tertiary_revisions <- complete(tertiary_revisions, "long", inc = TRUE)
tertiary_revisions\$Long_Los <- ifelse(tertiary_revisions\$Spell_Los > 5, 1, 0)
tertiary_revisions <- as.mids(tertiary_revisions)
#Exposure 1 - Distance Miles
library("Ime4")
Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for clustering m3.mi <- with(tertiary revisions, glm(Long Los ~ DistanceMiles + IMD score + HFRS Band +
sex + age_of_patient + infection + TV12mo + CV12mo + FinY +
family = "binomial"))
print(m3.mi)
Pool results across imputed datasets pooled_results <- pool(m3.mi)
Summarize pooled results with confidence intervals summary_pooled <- summary(pooled_results, conf.int = TRUE)

```
# Add Odds Ratios to the summary
summary pooled$OR <- exp(summary pooled$estimate)
summary pooled$Lower_Cl <- exp(summary_pooled$`2.5 %`)</pre>
summary_pooled$Upper_Cl <- exp(summary_pooled$`97.5 %`)</pre>
# Display the final table with Odds Ratios and Confidence Intervals
print(summary_pooled)
#check for evidence of multicollinearity?
library(car)
# Use the first imputed dataset for the VIF calculation
complete data <- complete(tertiary revisions, 1)
# Fit a logistic regression model on the complete dataset
vif model <- glm(Long Los ~ DistanceMiles + IMD score + HFRS Band +
          sex + age_of_patient + infection + TV12mo + CV12mo + FinY + ProvCode,
         data = complete data, family = "binomial")
# Calculate VIF
vif_values <- vif(vif_model)</pre>
print(vif_values)
#No evidence of multi-collinearity
#Is there evidence of non linearity?
# Custom function to add log-transformed variable and interaction term
add interaction <- function(data) {
 data$Log DistanceMiles <- log(data$DistanceMiles) # Add log-transformed variable
 data$Interaction <- data$DistanceMiles * data$Log DistanceMiles # Add interaction term
 return(data)
}
# Extract the long-format data including the original data
tertiary revisions modified <- complete(tertiary revisions, action = "long", include = TRUE)
# Apply the transformation to each imputed dataset
tertiary revisions modified <- do.call("rbind",
                     lapply(split(tertiary_revisions_modified,
tertiary revisions modified$.imp),
                         add interaction))
# Convert back to mids object
tertiary revisions modified <- as.mids(tertiary revisions modified)
```

BMJ Open

Fit the logistic regression model with the interaction term

model <- with(tertiary_revisions_modified, glm(Long_Los ~ DistanceMiles + Interaction, family = binomial(link = "logit"))) # Pool the results pooled_results <- pool(model)</pre> # Summarize pooled results summary_pooled <- summary(pooled_results, conf.int = TRUE)</pre> # Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"] # Print the p-value print(box_tidwell_p) #P value 0.002 Non linear # Load the required library library(splines) #AIC of non spline model model <- glm(Long_Los ~ DistanceMiles, data = tertiary_revisions, family = binomial) summary(model) #AIC 52853 # Define a function to fit and evaluate spline models with knots based on centiles evaluate centile splines <- function(centiles, data) { # Calculate knots based on the specified centiles knots <- quantile(data\$DistanceMiles, probs = centiles, na.rm = TRUE)</pre> # Fit a logistic regression model with natural splines using the calculated knots model spline <- $glm(Long Los \sim ns(DistanceMiles, knots = knots)$, family = binomial(link = "logit"), data = data) # Summarize the model

Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Summarize the model summary_model <- summary(model_spline)

Extract p-values for the spline terms
p_values <- summary_model\$coefficients[-1, "Pr(>|z|)"] # Exclude the intercept

Print the results

3	
4 5	
6	
7	
8	
9 10	
10	
12	
13	
14	
15	
17	
18	
19	
20	
22	
23	
24	
25 26	
27	
28	
29	
31	
32	
33	
34	
36	
37	
38	
39 40	
40 41	
42	
43	
44	
45 46	
47	
48	
49 50	
50	
52	
53	
54 55	
55 56	
57	
58	
59 60	
00	

cat("\nResults for centiles", centiles, ":\n") print(p_values)
<pre># Return the model and calculated knots for further inspection if needed return(list(model = model_spline, p_values = p_values, knots = knots)) }</pre>
Example centile configurations for 3, 4, and 5 knots centiles_3_knots <- c(0.05, 0.50, 0.95) # 5th, 50th, and 95th percentiles centiles_4_knots <- c(0.05, 0.35, 0.65, 0.95) # Custom centiles for 4 knots centiles_5_knots <- c(0.05, 0.25, 0.50, 0.75, 0.95) # 5 knots centiles
Evaluate models with centile-based knots using your dataset results_3_knots <- evaluate_centile_splines(centiles = centiles_3_knots, data = tertiary_revisions)
results_4_knots <- evaluate_centile_splines(centiles = centiles_4_knots, data =
tertiary_revisions) results_5_knots <- evaluate_centile_splines(centiles = centiles_5_knots, data = tertiary_revisions)
Compare models with centile-based knots cat("\nComparing models with different centile-based knots:\n") anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq")
<pre># Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots)</pre>
#52769, model with four knots best fit and improved fit from original linear model
#Run spline model with adjusted data excluding missing data library(splines)
<pre># For example, let's say you want 3 knots at specific percentiles knots <- quantile(tertiary_revisions\$DistanceMiles, probs = c(0.05, 0.35, 0.65, 0.95), na.rm = TRUE) print(knots)</pre>
spline_terms <- ns(tertiary_revisions\$DistanceMiles, knots = knots)

2	
3	model with custom splines <- glm(Long Los ~ ns(DistanceMiles, knots = knots) +
4	HERS Band + IMD score +
5	sor Lago of national TV/12mo LCV/12mo LEinV L DrovCodo
6	sex + age_or_patient + infection + 1v12mo + Cv12mo + Finv + ProvCode,
/	family = "binomial", data = tertiary_revisions)
8	
9	
10	summary(model with custom splines)
11	
12	#Constate a sequence of mean unit values for predicting
14	#Generate a sequence of mean unit values for predicting
15	
16	DistanceMiles_range <- seq(min(tertiary_revisions\$DistanceMiles),
17	max(tertiary_revisions\$DistanceMiles), length.out = 100)
18	
19	new data <- expand grid(
20	DistanceMiles - DistanceMiles, range
21	
22	sex = levels(tertiary_revisions\$sex), # Ensure it takes all factor levels
23	age_of_patient = mean(tertiary_revisions\$age_of_patient, na.rm = TRUE),
24	HFRS_Band = levels(tertiary_revisions\$HFRS_Band), # Ensuring correct factor levels
25	IMD score = mean(tertiary revisions\$IMD score, na.rm = TRUE),
26	FinY = levels(tertiary, revisionsSFinY) # Ensuring correct factor levels
27	CV12mo - mean(tertiary, revisions\$CV12mo, na rm - TRUE)
28	$\nabla I I I I O = III C I I I I I I I I I I I I I I I I$
29	IV12mo = mean(tertiary_revisions\$IV12mo, na.rm = IRUE),
3U 21	ProvCode = levels(tertiary_revisions\$ProvCode), # Ensuring correct factor levels
20	infection = levels(tertiary_revisions\$infection) # Ensuring correct factor levels
32	
34	
35	# Create a new dataset with a range of distances and miles and all other predictor variables
36	mercate a new dataset with a range of distances and miles and an other predictor variables
37	new_data <- expand.grid(Distanceivilles = Distanceivilles_range,
38	sex = unique(tertiary_revisions\$sex),
39	age_of_patient = mean(tertiary_revisions\$age_of_patient),
40	HFRS_Band = unique(tertiary_revisions\$HFRS_Band),
41	IMD score = mean(tertiary revisions\$IMD score),
42	FinY = unique(tertiary, revisions\$FinY)
43	CV(12mo = moon/tortiony, rovisions CV(12mo)
44	$CV12IIIO = Inteal((certiary_revisions)CV12IIIO),$
45	IV12mo = mean(tertiary_revisions\$IV12mo),
46	infection = unique(tertiary_revisions\$infection))
47	
48	
49 50	# Align the levels of ProvCode in new data to match the training data
51	new dataSProvCode <- factor(new dataSProvCode levels -
52	lovels(tortion, rovisions)
53	ieveis(tei tidry_revisions\$Provcoue))
54	
55	# Align the levels of all relevant categorical variables
56	new_data\$HFRS_Band <- factor(new_data\$HFRS_Band, levels =
57	levels(tertiary revisions\$HFRS Band))
58	new data\$sex <- factor(new data\$sex levels = levels(tertiary revisions\$sex))
59	$now_data \xi EinV < factor(now_data \xi EinV_love) = love) s(tortiony_revisions \xi EinV)$
60	new_uataşrını <- iactor(new_uataşrını, ieveis – ieveis(tertidiy_fevisions\$filit))

2	
3	
4	
5	
ć	
0	
7	
8	
9	
10	
10	
11	
12	
13	
14	
14	
15	
16	
17	
10	
18	
19	
20	
21	
27	
22	
23	
24	
25	
25	
20	
27	
28	
29	
20	
30	
31	
32	
33	
24	
34	
35	
36	
37	
20	
38	
39	
40	
⊿1	
40	
42	
43	
44	
15	
45	
46	
47	
48	
⊿0	
+7	
50	
51	
52	
52	
55	
54	
55	
56	
55	
5/	
58	
59	
60	

1

new_data\$infection <- factor(new_data\$infection, levels =
levels(tertiary_revisions\$infection))</pre>

#Factors are consistent with model

levels(new_data\$HFRS_Band) levels(tertiary_revisions\$HFRS_Band)

levels(new_data\$sex)
levels(tertiary_revisions\$sex)

levels(new_data\$FinY)
levels(tertiary_revisions\$FinY)

levels(new_data\$ProvCode) levels(tertiary_revisions\$ProvCode)

levels(new_data\$infection) levels(tertiary_revisions\$infection)

Check levels of ProvCode in both datasets
setdiff(levels(new_data\$ProvCode), levels(tertiary_revisions\$ProvCode)) # Levels in
new_data but not in tertiary_revisions
setdiff(levels(tertiary_revisions\$ProvCode), levels(new_data\$ProvCode)) # Levels in
tertiary_revisions but not in new_data

new_data\$ProvCode <- droplevels(new_data\$ProvCode)
Check for missing values in factor variables
sum(is.na(new_data\$ProvCode)) # Number of missing values in ProvCode</pre>

Ensure that ProvCode is a factor new_data\$ProvCode <- factor(new_data\$ProvCode, levels = levels(tertiary_revisions\$ProvCode))

Now try the prediction again
predicted_probs <- predict(model_with_custom_splines, newdata = new_data, type =
"response")</pre>

Combine mean_unit_range and predicted_probs into a data frame plot_data <- data.frame(DistanceMiles = DistanceMiles_range, predicted_prob = predicted_probs)

#Calculate 95% confidence intervals

4

5

6 7 8

9

10

11

12 13

14

15

16

17 18

19

20 21

22

23

24 25

26

27

28 29

36

37

38

39

40 41

42

43 44

45 46 47

48

49

50

51 52

53

54

55

56 57

```
# Obtain predicted values and standard errors for the new data
predictions <- predict(model with custom splines, newdata = new data, type = "link",
se.fit = TRUE)
# Calculate the confidence intervals for the log-odds scale (link scale)
# Use a 95% confidence level (z-value = 1.96 for a 95% CI)
z value <- 1.96
log odds lower <- predictions$fit - z value * predictions$se.fit
log_odds_upper <- predictions$fit + z_value * predictions$se.fit
# Convert the log-odds confidence intervals to probabilities
# First, apply the inverse link function (logistic function) to the log-odds
lower_prob <- plogis(log_odds_lower)</pre>
upper prob <- plogis(log odds upper)
# Combine the predicted probabilities and their confidence intervals into a data frame
plot data <- data.frame(
 DistanceMiles = new_data$DistanceMiles,
 predicted_prob = plogis(predictions$fit), # Logistic transformation of the link
 ci lower = lower prob,
 ci upper = upper prob
)
library(ggplot2)
# Plot the spline curve with confidence intervals
ggplot(plot data, aes(x = DistanceMiles)) +
 geom line(aes(y = predicted prob), color = "blue", size = 1) +
 geom_ribbon(aes(ymin = ci_lower, ymax = ci_upper), fill = "blue", alpha = 0.2) +
 labs(x = "Distance (Miles)", y = "Predicted Probability Readmission at 30 days") +
 theme minimal()
library(dplyr)
# Group by mean unit and calculate mean predicted prob and corresponding confidence
intervals
mean data <- plot data %>%
 group by(DistanceMiles) %>%
 summarise(
  mean_predicted_prob = mean(predicted_prob, na.rm = TRUE),
  mean ci lower = mean(ci lower, na.rm = TRUE),
  mean ci upper = mean(ci upper, na.rm = TRUE)
 )
```

BMJ Open

```
# Define specific breaks (e.g., 25, 50, 75, ..., up to the maximum)
breaks seq <- seq(0, max(mean data$DistanceMiles, na.rm = TRUE), by = 5)
library(ggplot2)
# Plot with specified increments on x-axis
ggplot(mean data, aes(x = DistanceMiles, y = mean predicted prob)) +
 geom point() + # Add points for mean predicted prob
 geom line() + # Connect points with a line
 geom_ribbon(aes(ymin = mean_ci_lower, ymax = mean_ci_upper), fill = "blue", alpha =
0.2) + # Add ribbon for confidence intervals
 labs(x = "Travel Distance (Miles)", y = "Mean Predicted Probability for Prolonged LOS", title
= "Spline curve predicted probability of prolonged LOS by patient travel distance") +
 scale x continuous(limits = c(0, max(mean data$DistanceMiles, na.rm = TRUE)), breaks =
breaks seq) +
 theme minimal() +
 theme(
  axis.title.x = element text(size = 14), # Increase x-axis title font size
  axis.title.y = element_text(size = 14), # Increase y-axis title font size
  axis.text.x = element text(size = 12), # Increase x-axis tick label font size
  axis.text.y = element_text(size = 12), # Increase y-axis tick label font size
  plot.title = element text(size = 16, hjust = 0.5) # Increase plot title font size and center it
 )
                                         reliew
#Exposure 2 - OffPeakDriveDistanceMiles
library("Ime4")
# Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for
clustering
m3.mi <- with(tertiary revisions, glm(Long Los ~ OffPeakDriveDistanceMiles + IMD score +
HFRS Band +
                      sex + age of patient + infection + TV12mo + CV12mo + FinY +
ProvCode,
                     family = "binomial"))
print(m3.mi)
# Pool results across imputed datasets
pooled results <- pool(m3.mi)
```

Summarize pooled results with confidence intervals summary_pooled <- summary(pooled_results, conf.int = TRUE)
Add Odds Ratios to the summary summary_pooled\$OR <- exp(summary_pooled\$estimate) summary_pooled\$Lower_Cl <- exp(summary_pooled\$`2.5 %`) summary_pooled\$Upper_Cl <- exp(summary_pooled\$`97.5 %`)
Display the final table with Odds Ratios and Confidence Intervals print(summary_pooled)
#check for evidence of multicollinearity?
library(car)
Use the first imputed dataset for the VIF calculation complete_data <- complete(tertiary_revisions, 1)
<pre># Fit a logistic regression model on the complete dataset vif_model <- glm(Read30days ~ DistanceMiles + IMD_score + HFRS_Band +</pre>
<pre># Calculate VIF vif_values <- vif(vif_model) print(vif_values)</pre>
#No evidence of multi-collinearity
#Is there evidence of non linearity?
<pre># Custom function to add log-transformed variable and interaction term add_interaction <- function(data) { data\$Log_OffPeakDriveDistanceMiles <- log(data\$OffPeakDriveDistanceMiles) # Add log- transformed variable data\$Interaction <- data\$OffPeakDriveDistanceMiles * data\$Log_OffPeakDriveDistanceMiles # Add interaction term return(data) }</pre>
Extract the long-format data including the original data tertiary_revisions_modified <- complete(tertiary_revisions, action = "long", include = TRUE)
Apply the transformation to each imputed dataset tertiary_revisions_modified <- do.call("rbind",

BMJ Open

lapply(split(tertiary_revisions_modified, tertiary revisions modified\$.imp), add interaction)) # Convert back to mids object tertiary revisions modified <- as.mids(tertiary revisions modified) # Fit the logistic regression model with the interaction term model <- with(tertiary_revisions_modified, glm(Long_Los ~ OffPeakDriveDistanceMiles + Interaction, family = binomial(link = "logit"))) # Pool the results pooled_results <- pool(model)</pre> # Summarize pooled results summary pooled <- summary(pooled results, conf.int = TRUE) # Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"] # Print the p-value elie print(box_tidwell_p) #0.003 #AIC of non spline model model <- glm(Long Los ~ OffPeakDriveDistanceMiles, data = tertiary revisions, family = binomial) summary(model) #AIC 52853 # Define a function to fit and evaluate spline models with knots based on centiles evaluate centile splines <- function(centiles, data) { # Calculate knots based on the specified centiles knots <- quantile(data\$OffPeakDriveDistanceMiles, probs = centiles, na.rm = TRUE) # Fit a logistic regression model with natural splines using the calculated knots model spline <- glm(Long Los ~ ns(OffPeakDriveDistanceMiles, knots = knots), family = binomial(link = "logit"), data = data) # Summarize the model summary model <- summary(model spline)

3	# Extract p-values for the spline terms
4	n values $<$ summary models coefficients [-1 "Pr(> z)"] # Exclude the intercent
5	
6	
7	# Print the results
8	cat("\nResults for centiles" centiles "·\n")
9	erint(n velues)
10	print(p_values)
11	
12	# Return the model and calculated knots for further inspection if needed
13	return/list(model - model spling p values - p values knots - knots))
14	$\frac{1}{2} = \frac{1}{2} = \frac{1}$
15	}
16	
10	# Example centile configurations for 3. 4. and 5 knots
17	contiles 3 knots $< c(0.05, 0.50, 0.05)$ # 5th 50th and 05th percentiles
10	$\frac{1}{1000} = \frac{1}{1000} = 1$
19	centiles_4_knots <- c(0.05, 0.35, 0.65, 0.95) # Custom centiles for 4 knots
20	centiles 5 knots <- c(0.05, 0.25, 0.50, 0.75, 0.95) # 5 knots centiles
21	
22	# Evaluate models with contile based knots using your detect
23	# Evaluate models with centile-based knots using your dataset
24	results_3_knots <- evaluate_centile_splines(centiles = centiles_3_knots, data =
25	tertiary revisions)
26	results 4 knots <- evaluate centile solines(centiles = centiles 4 knots data =
27	testions revisions)
28	tertiary_revisions)
29	results_5_knots <- evaluate_centile_splines(centiles = centiles_5_knots, data =
30	tertiary revisions)
31	
32	
33	# Compare models with centile-based knots
34	cat("\nComparing models with different centile-based knots:\n")
25	
22	anova(results 3 knotsSmodel, results 4 knotsSmodel, results 5 knotsSmodel, test =
36	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test =
36 37	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq")
36 37 38	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq")
35 36 37 38 39	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model</pre>
36 37 38 39 40	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n")</pre>
35 36 37 38 39 40 41	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots)</pre>
33 36 37 38 39 40 41 42	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots)</pre>
35 36 37 38 39 40 41 42 43	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n")</pre>
33 36 37 38 39 40 41 42 43 44	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results 4 knots\$knots)
33 36 37 38 39 40 41 42 43 44	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n")
33 36 37 38 39 40 41 42 43 44 45 46	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots)
33 36 37 38 39 40 41 42 43 44 45 46 47	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots)
33 36 37 38 39 40 41 42 43 44 45 46 47 40	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots)
33 36 37 38 39 40 41 42 43 44 45 46 47 48	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms</pre>
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms</pre>
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms #Dues enline medial with a diversed deline and diversed by </pre>
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms #Run spline model with adjusted data excluding missing data</pre>
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms #Run spline model with adjusted data excluding missing data library(splines)</pre>
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms #Run spline model with adjusted data excluding missing data library(splines) # For example, let's say you want 3 knots at specific percentiles</pre>
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms #Run spline model with adjusted data excluding missing data library(splines) # For example, let's say you want 3 knots at specific percentiles knots <- quantile(tertiary, revisions\$OffPookDriveDistanceMilos, probs = c(0.05, 0.25, 0.65)
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms #Run spline model with adjusted data excluding missing data library(splines) # For example, let's say you want 3 knots at specific percentiles knots <- quantile(tertiary_revisions\$OffPeakDriveDistanceMiles, probs = c(0.05, 0.35, 0.65, 0.05)</pre>
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms #Run spline model with adjusted data excluding missing data library(splines) # For example, let's say you want 3 knots at specific percentiles knots <- quantile(tertiary_revisions\$OffPeakDriveDistanceMiles, probs = c(0.05, 0.35, 0.65, 0.95), na.rm = TRUE)
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57	<pre>anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms #Run spline model with adjusted data excluding missing data library(splines) # For example, let's say you want 3 knots at specific percentiles knots <- quantile(tertiary_revisions\$OffPeakDriveDistanceMiles, probs = c(0.05, 0.35, 0.65, 0.95), na.rm = TRUE) print(knots)</pre>
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms #Run spline model with adjusted data excluding missing data library(splines) # For example, let's say you want 3 knots at specific percentiles knots <- quantile(tertiary_revisions\$OffPeakDriveDistanceMiles, probs = c(0.05, 0.35, 0.65, 0.95), na.rm = TRUE) print(knots)
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms #Run spline model with adjusted data excluding missing data library(splines) # For example, let's say you want 3 knots at specific percentiles knots <- quantile(tertiary_revisions\$OffPeakDriveDistanceMiles, probs = c(0.05, 0.35, 0.65, 0.95), na.rm = TRUE) print(knots) spline_terms <- ns(tertiary_revisions\$OffPeakDriveDistanceMiles_knots = knots)
33 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60	anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq") # Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots) #52718, model with four knots best fit and significant spline terms #Run spline model with adjusted data excluding missing data library(splines) # For example, let's say you want 3 knots at specific percentiles knots <- quantile(tertiary_revisions\$OffPeakDriveDistanceMiles, knots = knots) spline_terms <- ns(tertiary_revisions\$OffPeakDriveDistanceMiles, knots = knots)

```
model with custom splines <- glm(Long Los ~ ns(OffPeakDriveDistanceMiles, knots =
knots) + HFRS Band + IMD score +
                   sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode,
                  family = "binomial", data = tertiary_revisions)
summary(model_with_custom_splines)
#Generate a sequence of mean unit values for predicting
DistanceMiles_range <- seq(min(tertiary_revisions$OffPeakDriveDistanceMiles),
max(tertiary revisions$OffPeakDriveDistanceMiles), length.out = 100)
new data <- expand.grid(
 OffPeakDriveDistanceMiles = DistanceMiles range,
 sex = levels(tertiary revisions$sex), # Ensure it takes all factor levels
 age of patient = mean(tertiary revisions$age of patient, na.rm = TRUE),
 HFRS Band = levels(tertiary revisions$HFRS Band), # Ensuring correct factor levels
 IMD_score = mean(tertiary_revisions$IMD_score, na.rm = TRUE),
 FinY = levels(tertiary revisions$FinY), # Ensuring correct factor levels
 CV12mo = mean(tertiary_revisions$CV12mo, na.rm = TRUE),
 TV12mo = mean(tertiary revisions$TV12mo, na.rm = TRUE),
 ProvCode = levels(tertiary revisions$ProvCode), # Ensuring correct factor levels
 infection = levels(tertiary_revisions$infection) # Ensuring correct factor levels
)
# Create a new dataset with a range of distances and miles and all other predictor variables
new_data <- expand.grid(DistanceMiles = DistanceMiles_range,
             sex = unique(tertiary revisions$sex),
             age of patient = mean(tertiary revisions$age of patient),
             HFRS Band = unique(tertiary revisions$HFRS Band),
             IMD score = mean(tertiary revisions$IMD score),
             FinY = unique(tertiary_revisions$FinY),
             CV12mo = mean(tertiary revisions$CV12mo),
             TV12mo = mean(tertiary revisions$TV12mo),
             infection = unique(tertiary_revisions$infection))
# Align the levels of ProvCode in new data to match the training data
new data$ProvCode <- factor(new data$ProvCode, levels =
levels(tertiary_revisions$ProvCode))
# Align the levels of all relevant categorical variables
```

60

2	
3	new_data\$HFRS_Band <- factor(new_data\$HFRS_Band_levels =
4	lovels(tertiany, revisions\$HEPS, Band))
5	revers(ter tiary_revisions) in KS_band))
6	new_data\$sex <- factor(new_data\$sex, levels = levels(tertiary_revisions\$sex))
7	new_data\$FinY <- factor(new_data\$FinY, levels = levels(tertiary_revisions\$FinY))
8	new data\$infection <- factor(new data\$infection, levels =
9	levels(tertiary revisionsSinfection))
10	
11	#Factors are consistent with model
12	#Factors are consistent with model
13	
14	levels(new_data\$HFRS_Band)
15	levels(tertiary revisions\$HFRS Band)
10	, ,_ ,_ ,
17	levels(new data\$ser)
10	levels(new_data;sex)
20	levels(tertiary_revisions\$sex)
20	
22	levels(new_data\$FinY)
23	levels(tertiary revisions\$FinY)
24	
25	lovels(now, data\$BroyCode)
26	levels(new_ualasProveoue)
27	levels(tertiary_revisions\$ProvCode)
28	
29	levels(new_data\$infection)
30	levels(tertiary revisionsSinfection)
31	
32	# Charly loyals of DroyCode in both detects
33	
34	setdiff(levels(new_data\$ProvCode), levels(tertiary_revisions\$ProvCode)) # Levels in
35	new_data but not in tertiary_revisions
36	setdiff(levels(tertiary_revisions\$ProvCode), levels(new_data\$ProvCode)) # Levels in
37	tertiary revisions but not in new data
38	
39	now dataś ProvCada z dranlovalc/now dataś ProvCada)
40	
41	# Check for missing values in factor variables
42	sum(is.na(new_data\$ProvCode)) # Number of missing values in ProvCode
44	
45	# Ensure that ProvCode is a factor
46	new data\$ProvCode <- factor(new data\$ProvCode levels =
47	lovels/tertiany revisions(ProvCode))
48	levels(tertiary_revisions\$ProvCode))
49	
50	# Now try the prediction again
51	predicted probs <- predict(model with custom splines, newdata = new data, type =
52	"response")
53	
54	
55	
56	
57	
58	# Combine mean unit range and predicted probs into a data frame
59	

2	
3	
4	
5	
6	
7	
/	
8	
9	
10	
11	
12	
12	
13	
14	
15	
16	
17	
18	
10	
19	
20	
21	
22	
23	
24	
27	
25	
26	
27	
28	
29	
30	
21	
21	
32	
33	
34	
35	
36	
37	
20	
38	
39	
40	
41	
42	
43	
11	
44 15	
45	
46	
47	
48	
49	
50	
50 E 1	
21	
52	
53	
54	
55	
56	
57	
57	
58	
59	

60

1

plot_data <- data.frame(OffPeakDriveDistanceMiles = DistanceMiles_range, predicted_prob = predicted probs) #Calculate 95% confidence intervals # Obtain predicted values and standard errors for the new data predictions <- predict(model_with_custom_splines, newdata = new_data, type = "link", se.fit = TRUE) # Calculate the confidence intervals for the log-odds scale (link scale) # Use a 95% confidence level (z-value = 1.96 for a 95% CI) z value <- 1.96 log odds lower <- predictions\$fit - z value * predictions\$se.fit log odds upper <- predictions\$fit + z value * predictions\$se.fit # Convert the log-odds confidence intervals to probabilities # First, apply the inverse link function (logistic function) to the log-odds lower_prob <- plogis(log_odds_lower)</pre> upper_prob <- plogis(log_odds_upper)</pre> # Combine the predicted probabilities and their confidence intervals into a data frame plot_data <- data.frame(</pre> DistanceMiles = new data\$OffPeakDriveDistanceMiles, predicted_prob = plogis(predictions\$fit), # Logistic transformation of the link ci lower = lower prob, ci upper = upper prob) library(ggplot2) # Plot the spline curve with confidence intervals ggplot(plot data, aes(x = DistanceMiles)) + geom line(aes(y = predicted prob), color = "blue", size = 1) + geom_ribbon(aes(ymin = ci_lower, ymax = ci_upper), fill = "blue", alpha = 0.2) + labs(x = "Distance (Miles)", y = "Predicted Probability Readmission at 30 days") + theme minimal() library(dplyr) # Group by mean_unit and calculate mean predicted_prob and corresponding confidence intervals mean data <- plot data %>% group by(DistanceMiles) %>% summarise(mean predicted prob = mean(predicted prob, na.rm = TRUE),

```
mean ci lower = mean(ci lower, na.rm = TRUE),
               mean ci upper = mean(ci upper, na.rm = TRUE)
              )
             # Define specific breaks (e.g., 25, 50, 75, ..., up to the maximum)
10
             breaks_seq <- seq(0, max(mean_data$DistanceMiles, na.rm = TRUE), by = 5)</pre>
11
12
13
             library(ggplot2)
14
             # Plot with specified increments on x-axis
15
             ggplot(mean data, aes(x = DistanceMiles, y = mean predicted prob)) +
16
              geom point() + # Add points for mean predicted prob
17
18
              geom line() + # Connect points with a line
19
              geom ribbon(aes(ymin = mean ci lower, ymax = mean ci upper), fill = "blue", alpha =
20
             0.2) + # Add ribbon for confidence intervals
21
              labs(x = "Off Peak Drive Distance Miles", y = "Mean Predicted Probability for Prolonged
22
             LOS", title = "Spline curve predicted probability of prolonged LOS by patient driving
23
24
             distance") +
25
              scale x continuous(limits = c(0, max(mean data$DistanceMiles, na.rm = TRUE)), breaks =
26
             breaks seq) +
27
              theme minimal() +
28
29
              theme(
30
               axis.title.x = element text(size = 14), # Increase x-axis title font size
31
               axis.title.y = element_text(size = 14), # Increase y-axis title font size
32
               axis.text.x = element_text(size = 12), # Increase x-axis tick label font size
33
               axis.text.y = element text(size = 12), # Increase y-axis tick label font size
34
35
               plot.title = element_text(size = 16, hjust = 0.5) # Increase plot title font size and center it
36
              )
37
38
39
40
41
42
             #Exposure 3 - PeakDriveTime
43
44
45
46
47
             library("Ime4")
48
49
             # Fit logistic regression on imputed datasets include ProvCode in fixed effects to account for
50
             clustering
51
52
             m3.mi <- with(tertiary revisions, glm(Long Los ~ PeakDriveTime + IMD score + HFRS Band
53
             +
54
                                    sex + age of patient + infection + TV12mo + CV12mo + FinY +
55
             ProvCode,
56
57
                                   family = "binomial"))
58
59
60
```

3	
5	
4	
5	
6	
7	
ç	
0	
9	
10	
11	
12	
12	
15	
14	
15	
16	
17	
10	
10	
19	
20	
21	
22	
22	
23	
24	
25	
26	
27	
20	
20	
29	
30	
31	
32	
22	
27	
34	
35	
36	
37	
38	
20	
39	
40	
41	
42	
43	
11	
44	
45	
46	
47	
48	
10	
77	
50	
51	
52	
53	
51	
54	
55	
56	
57	
58	
50	
55	
60	

1 2

print(m3.mi)

Pool results across imputed datasets pooled results <- pool(m3.mi)</pre> # Summarize pooled results with confidence intervals summary pooled <- summary(pooled results, conf.int = TRUE) # Add Odds Ratios to the summary summary pooled\$OR <- exp(summary pooled\$estimate) summary pooled\$Lower Cl <- exp(summary pooled\$`2.5 %`) summary_pooled\$Upper_Cl <- exp(summary_pooled\$`97.5 %`)</pre> # Display the final table with Odds Ratios and Confidence Intervals print(summary_pooled) #check for evidence of multicollinearity? library(car) # Use the first imputed dataset for the VIF calculation complete data <- complete(tertiary revisions, 1) # Fit a logistic regression model on the complete dataset vif model <- glm(Read30days ~ DistanceMiles + IMD score + HFRS Band + sex + age of patient + infection + TV12mo + CV12mo + FinY + ProvCode, data = complete data, family = "binomial") # Calculate VIF vif_values <- vif(vif_model)</pre> print(vif values) #Is there evidence of non linearity? # Custom function to add log-transformed variable and interaction term add interaction <- function(data) { data\$Log PeakDriveTime <- log(data\$PeakDriveTime) # Add log-transformed variable data\$Interaction <- data\$PeakDriveTime * data\$Log_PeakDriveTime # Add interaction term return(data) } # Extract the long-format data including the original data tertiary_revisions_modified <- complete(tertiary_revisions, action = "long", include = TRUE)</pre>

<pre># Apply the transformation to each imputed dataset tertiary_revisions_modified <- do.call("rbind",</pre>
Convert back to mids object tertiary_revisions_modified <- as.mids(tertiary_revisions_modified)
Fit the logistic regression model with the interaction term model <- with(tertiary_revisions_modified, glm(Long_Los ~ PeakDriveTime + Interaction, family = binomial(link = "logit")))
Pool the results pooled_results <- pool(model)
Summarize pooled results summary_pooled <- summary(pooled_results, conf.int = TRUE)
Extract the p-value for the interaction term box_tidwell_p <- summary_pooled[summary_pooled\$term == "Interaction", "p.value"]
Print the p-value print(box_tidwell_p)
#P value 0.000916
#AIC of non spline model
model <- glm(Long_Los ~ PeakDriveTime, data = tertiary_revisions, family = binomial) summary(model)
#AIC 52843
<pre># Define a function to fit and evaluate spline models with knots based on centiles evaluate_centile_splines <- function(centiles, data) { # Calculate knots based on the specified centiles knots <- quantile(data\$PeakDriveTime, probs = centiles, na.rm = TRUE)</pre>
Fit a logistic regression model with natural splines using the calculated knots model_spline <- glm(Long_Los ~ ns(PeakDriveTime, knots = knots), family = binomial(link = "logit"), data = data)
Summarize the model summary_model <- summary(model_spline)

Extract p-values for the spline terms p_values <- summary_model\$coefficients[-1, "Pr(> z)"] # Exclude the intercept
Print the results cat("\nResults for centiles", centiles, ":\n") print(p_values)
<pre># Return the model and calculated knots for further inspection if needed return(list(model = model_spline, p_values = p_values, knots = knots)) }</pre>
Example centile configurations for 3, 4, and 5 knots centiles_3_knots <- c(0.05, 0.50, 0.95) # 5th, 50th, and 95th percentiles centiles_4_knots <- c(0.05, 0.35, 0.65, 0.95) # Custom centiles for 4 knots centiles_5_knots <- c(0.05, 0.25, 0.50, 0.75, 0.95) # 5 knots centiles
<pre># Evaluate models with centile-based knots using your dataset results_3_knots <- evaluate_centile_splines(centiles = centiles_3_knots, data = tertiary_revisions) results_4_knots <- evaluate_centile_splines(centiles = centiles_4_knots, data = tertiary_revisions) results_5_knots <- evaluate_centile_splines(centiles = centiles_5_knots, data = tertiary_revisions)</pre>
<pre># Compare models with centile-based knots cat("\nComparing models with different centile-based knots:\n") anova(results_3_knots\$model, results_4_knots\$model, results_5_knots\$model, test = "Chisq")</pre>
<pre># Print the calculated knot locations for each model cat("\nKnot locations for 3 knots:\n") print(results_3_knots\$knots) cat("\nKnot locations for 4 knots:\n") print(results_4_knots\$knots) cat("\nKnot locations for 5 knots:\n") print(results_5_knots\$knots)</pre>
#52715, model with four knots best fit and significant spline terms and most parsimonious
<pre>#Run spline model with adjusted data excluding missing data library(splines) # For example, let's say you want 3 knots at specific percentiles knots <- quantile(tertiary_revisions\$PeakDriveTime, probs = c(0.05, 0.35, 0.65, 0.95), na.rm = TRUE) print(knots)</pre>
<pre>spline_terms <- ns(tertiary_revisions\$PeakDriveTime, knots = knots)</pre>

model_with_custom_splines <- glm(Long_Los ~ ns(PeakDriveTime, knots = HFRS_Band + IMD_score + sex + age_of_patient + infection + TV12mo + CV12mo + family = "binomial", data = tertiary_revisions)	knots) + FinY + ProvCode,
summary(model_with_custom_splines)	
#Generate a sequence of mean unit values for predicting	
DistanceMiles_range <- seq(min(tertiary_revisions\$PeakDriveTime), max(tertiary_revisions\$PeakDriveTime), length.out = 100)	
<pre>new_data <- expand.grid(PeakDriveTime = DistanceMiles_range, sex = levels(tertiary_revisions\$sex), # Ensure it takes all factor levels age_of_patient = mean(tertiary_revisions\$age_of_patient, na.rm = TRUE), HFRS_Band = levels(tertiary_revisions\$HFRS_Band), # Ensuring correct fact IMD_score = mean(tertiary_revisions\$IMD_score, na.rm = TRUE), FinY = levels(tertiary_revisions\$FinY), # Ensuring correct factor levels CV12mo = mean(tertiary_revisions\$CV12mo, na.rm = TRUE), TV12mo = mean(tertiary_revisions\$TV12mo, na.rm = TRUE), ProvCode = levels(tertiary_revisions\$ProvCode), # Ensuring correct factor infection = levels(tertiary_revisions\$infection) # Ensuring correct factor levels)</pre>	ctor levels levels evels
# Align the levels of ProvCode in new_data to match the training data new_data\$ProvCode <- factor(new_data\$ProvCode, levels = levels(tertiary_revisions\$ProvCode))	
<pre># Align the levels of all relevant categorical variables new_data\$HFRS_Band <- factor(new_data\$HFRS_Band, levels = levels(tertiary_revisions\$HFRS_Band)) new_data\$sex <- factor(new_data\$sex, levels = levels(tertiary_revisions\$se new_data\$FinY <- factor(new_data\$FinY, levels = levels(tertiary_revisions\$ new_data\$infection <- factor(new_data\$infection, levels = levels(tertiary_revisions\$infection))</pre>	ex)) FinY))
#Factors are consistent with model	
levels(new_data\$HFRS_Band) levels(tertiary_revisions\$HFRS_Band)	
levels(new_data\$sex)	
Enseignement Superieur (ABES) Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

2	
3	
4	ieveis(tertiary_revisions\$sex)
5	
6	levels(new data\$FinY)
7	levels(tertiary revisions\$FinY)
, 8	
9	
10	levels(new_data\$ProvCode)
10	levels(tertiary revisions\$ProvCode)
17	· · · · · ·
12	lough/nour doto(infoction)
13	levels(new_data\$infection)
14	levels(tertiary_revisions\$infection)
15	
10	# Check levels of ProvCode in both datasets
17	weneek levels of Provedue in both dutusets
18	setdin(levels(new_datasprovCode), levels(tertiary_revisions\$provCode)) # Levels in
19	new_data but not in tertiary_revisions
20	setdiff(levels(tertiary revisions\$ProvCode), levels(new data\$ProvCode)) # Levels in
21	tertiary revisions but not in new data
22	
23	
24	new_data\$ProvCode <- droplevels(new_data\$ProvCode)
25	# Check for missing values in factor variables
26	sum(is.na(new_data\$ProvCode)) # Number of missing values in ProvCode
2/	
28	
29	# Ensure that ProvLode is a factor
30	new_data\$ProvCode <- factor(new_data\$ProvCode, levels =
31	levels(tertiary_revisions\$ProvCode))
32	
33	# Now try the prediction again
34 25	
35	predicted_probs <- predict(model_with_custom_splines, newdata = new_data, type =
30 27	"response")
20	
30	
39	
40	
41	
42	# Combine mean unit range and predicted probs into a data frame
43	plot_data <- data.frame(PeakDriveTime = DistanceMiles_range, predicted_prob =
44	prod_cata probe)
45	predicted_probs)
46	
47	#Calculate 95% confidence intervals
48	
49	# Obtain predicted values and standard errors for the new data
50	
51	predictions <- predict(model_with_custom_splines, newdata = new_data, type = "link",
52	se.fit = TRUE)
53	
54	# Calculate the confidence intervals for the log-odds scale (link scale)
55	# Lies a OEV confidence level (a value = 4.00 for a OEV CI)
56	# Use a 95% confidence level (z-value = 1.96 for a 95% CI)
57	z_value <- 1.96
58	log_odds_lower <- predictions\$fit - z_value * predictions\$se.fit
59	log_odds_upper <- predictions\$fit + z_value * predictions\$se fit
60	

1

BMJ Open

```
# Convert the log-odds confidence intervals to probabilities
# First, apply the inverse link function (logistic function) to the log-odds
lower prob <- plogis(log odds lower)</pre>
upper_prob <- plogis(log_odds_upper)</pre>
# Combine the predicted probabilities and their confidence intervals into a data frame
plot data <- data.frame(
 DriveTime = new data$PeakDriveTime,
 predicted_prob = plogis(predictions$fit), # Logistic transformation of the link
 ci lower = lower prob,
 ci upper = upper prob
)
library(ggplot2)
# Plot the spline curve with confidence intervals
ggplot(plot data, aes(x = DriveTime)) +
 geom line(aes(y = predicted prob), color = "blue", size = 1) +
 geom_ribbon(aes(ymin = ci_lower, ymax = ci_upper), fill = "blue", alpha = 0.2) +
 labs(x = "Distance (Miles)", y = "Predicted Probability Readmission at 30 days") +
 theme_minimal()
library(dplyr)
# Group by mean unit and calculate mean predicted prob and corresponding confidence
intervals
mean data <- plot data %>%
 group_by(DriveTime) %>%
 summarise(
  mean predicted prob = mean(predicted prob, na.rm = TRUE),
  mean ci lower = mean(ci lower, na.rm = TRUE),
  mean ci upper = mean(ci upper, na.rm = TRUE)
 )
# Define specific breaks (e.g., 25, 50, 75, ..., up to the maximum)
breaks seq <- seq(0, max(mean data$DriveTime, na.rm = TRUE), by = 5)
library(ggplot2)
# Plot with specified increments on x-axis
ggplot(mean_data, aes(x = DriveTime, y = mean_predicted_prob)) +
 geom point() + # Add points for mean predicted prob
 geom line() + # Connect points with a line
```

```
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
```

```
geom_ribbon(aes(ymin = mean_ci_lower, ymax = mean_ci_upper), fill = "blue", alpha =
0.2) + # Add ribbon for confidence intervals
 labs(x = "Peak Drive Times (Minutes)", y = "Mean Predicted Probability for Prolonged LOS",
title = "Spline curve predicted probability of prolonged LOS by patient driving times") +
 scale x continuous(limits = c(0, max(mean data$DriveTime, na.rm = TRUE)), breaks =
breaks seq) +
 theme minimal() +
 theme(
            = 1.
.e = 14),
.ize = 12), *
.(size = 12), # h
.(size = 16, hjust = 0
  axis.title.x = element_text(size = 14), # Increase x-axis title font size
  axis.title.y = element text(size = 14), # Increase y-axis title font size
  axis.text.x = element text(size = 12), # Increase x-axis tick label font size
  axis.text.y = element text(size = 12), # Increase y-axis tick label font size
  plot.title = element_text(size = 16, hjust = 0.5) # Increase plot title font size and center it
 )
```

####END####

infection

Spell_Los

OffPeakDriveDistanceMiles

008808.

TV12mo

PeakDriveTime

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml