Study Title: A multi-centre validation of the ability of Suspicion Of Sepsis (SOS) diagnosis codes to identify bacterial infections diagnosed during hospital inpatient spells: a review of medical records.

Internal Reference No: PHT/2018/56

IRAS Project ID: 253598

Date and Version No: Version 1.0 November 1st, 2018

Chief Investigator: Paul Schmidt, Consultant, Portsmouth Hospitals NHS Trust (PHT)

Investigators: Matt Inada-Kim, Consultant, Hampshire Hospitals NHS Foundation Trust

Isabel Gonzalez, Consultant, South Tees Hospitals NHS Foundation Trust

Emmanuel Nsutebu, Consultant, Royal Liverpool and Broadgreen University Hospitals Trust

Kobus Preller, Consultant, Cambridge University Hospitals NHS Trust

Gary Smith, member of Clinical Outcomes Research Group at Portsmouth Hospitals NHS Trust

Paul Meredith, Senior Data Analyst, Portsmouth Hospitals NHS Trust (PHT)

Carole Fogg, Senior Lecturer, Portsmouth Hospitals NHS Trust (PHT)/University of Portsmouth

Sponsor: Portsmouth Hospitals NHS Trust

Funder: NHS Improvement

Signatures:

Dr P E Schmidt

TABLE OF CONTENTS

1. AMENDMENT HISTORY

Amendment No.	Protocol Version No.	Date issued	Author(s) of changes	Details of Changes made
0	1.0	1/11/2018	PS	Version submitted on IRAS

2. SYNOPSIS

Study Title	A multi-centre validation of the ability of Suspicion Of Sepsis (SOS) diagnosis codes to identify bacterial infections diagnosed during hospital inpatient spells: a review of medical records.
Internal ref. no.	
Problem statement	The accuracy of clinical coding in reflecting clinical diagnosis of bacterial infection in hospital is unknown.
Research question / hypothesis	Does the classification of hospital spells by Suspicion of Sepsis (SOS) diagnosis codes at diascharge accurately reflect clinical conviction of bacterial infection?
Study Design	Retrospective multi-centre observational study consisting of clinical notes review on each site by two clinical reviewers to determine whether a patient experienced a bacterial infection during hospitalisation. Blinded comparison to diagnosis coding at discharge for the presence of bacterial infection.
Study Participants	Medical records of adult hospitalised patients admitted between April 1st, 2017 and March 31st, 2018.
Planned Sample Size	720 notes reviews; 5 NHS hospitals will each contribute 144 reviews
Follow-up duration	Not applicable.
Planned Study Period	6 months
Primary Objective	To estimate the accuracy of clinical coding at hospital discharge in capturing episodes of bacterial infection in hospital admissions, assessed against a gold standard of structured clinical notes review.
Secondary Objectives	1. To estimate the prevalence of bacterial infection in the whole sample. 2. To describe variation in practice of clinical coding for bacterial infection between hospitals. 3. To assess the quality of evidence that the clinical conviction of the treating team for the presence of bacterial infection is based on. 4. To describe differences in demographics and clinical characteristics between SOS and non-SOS cohorts summarised by the structured case note reviews
Primary Endpoint	Diagnostic accuracy of the SOS classification (proportion of true positives, true negatives, false positives and false negatives of the clinical coding as compared to the case-note review)
Secondary Endpoints	1. Overall prevalence of bacterial infection (as per case-note reviews). 2. Measurement of variation in coding accuracy between sites a. variation in diagnostic accuracy. b. The distribution and variation in false positive rates and false negative rates across the sites as an assessment of systematic error in coding practices. 3. Quality of evidence for bacterial infection a. The proportions of patients with categorised clinical and laboratory evidence for bacterial infection as per case note review b. The proportions of patients categorised as SOS by level of clinical conviction of the presence of bacterial infection 4. Summary statistics of population characteristics, treatment and outcome differences between SOS and non-SOS cohorts as defined by case-note review?.
Intervention (s)	None

3. ABBREVIATIONS

CQUIN	Commissioning for Quality and Innovation
CSSPB	Cross-System Sepsis Programme board
FCE	Finished Consultant Episode
ICD	International Classification of Diseases
NEWS	National Early Warning Score
NHS	National Health Service
NHSE	NHS England
NHSI	NHS Improvement
PAS	Patient Administration System
PSMU	Patient Safety Management Unit
SOS	Suspicion of Sepsis

4. LAY SUMMARY

The Suspicion of Sepsis group (SOS) is a bundle of 460 ICD-10 (International Classification of Diseases Version 10) diagnosis codes used for all possible bacterial infections. It was created to better measure the incidence and outcomes from infection and sepsis, and whether national quality improvement projects are improving outcomes from sepsis. Using information from hospital admissions that are already reported avoids costly data collection.

The use of hospital diagnosis codes for this purpose has not yet been validated, meaning it is not known how accurately it indicates that the patient had a bacterial infection. In part, validation has to measure how accurate clinical coders are at interpreting medical notes which the medical team made during the patient's illness. These medical notes might mention the right diagnosis, or it might only mention the clinical and laboratory data supporting a diagnosis of bacterial infection. Sometimes doctors might also use terms such as "possible" or probable" to indicate uncertainty, which can then cause difficulty for anyone trying to understand exactly how to code an admission. NHS Improvement, under the direction of Celia Ingham-Clark, Interim Director for Patient Safety and NHS England Medical Director for Clinical Effectiveness, agreed to fund a study to investigate this relationship.

We are proposing a study at five NHS Trusts in England to compare the accuracy of clinical coder's coding of diagnoses using the medical notes and discharge summaries, to blinded structured case note reviews by two medical reviewers. Their assessment of the evidence supporting the presence or absence of bacterial infection will be the best available standard, as they will be using all sources of information, namely vital signs charts, laboratory data, imaging reports, medication and fluid charts and medical notes.

5. BACKGROUND AND RATIONALE

It has been difficult to ascertain the size of the burden of infective diagnoses admitted to English hospitals. Without this, progress to determine national outcomes performance, develop improvement strategies and benchmark organisations in the management of infection and sepsis, antibiotic stewardship, antimicrobial resistance and all-cause deterioration has been difficult.

The Suspicion of Sepsis group (SOS) comprises a bundle of ICD-10 codes pertaining to bacterial infection (*Defining and Measuring Suspicion of Sepsis- an analysis of routine data; Inada-Kim et al; BMJ Open 2017*)[1]. It was developed as a solution to the measurement challenge by utilising routinely reported hospital episode statistics data as it was quickly recognised that asking organisations to capture additional data would not be possible.

A pilot study in Queen Alexandra Hospital, Portsmouth by the lead investigators has demonstrated that 86% of positive bacteraemias due to E. Coli and Staph. Aureus over a 2-year period fell within episodes coded with the SOS codes (unpublished data). Another analysis of 242,000 admissions revealed that SOS coded episodes has at least 5-fold odds of death, at least 4-fold odds of ICU admission, and at least 8-fold odds of an ICU stay ≥ 3days compared to non-SOS coded episodes[2].

NHS England and Digital explored the use of SOS at a national level from 2010-2017, compiled a report and presented their findings to the Cross-System Sepsis Programme board (CSSPB) in June 2017. It has shown that SOS admissions have become more prevalent since 2010, to encompass 35% of all emergency admissions in England (PMSU data) from a baseline of 25% in 2010. Following this, Celia Ingham-Clark, Medical Director for Clinical Effectiveness NHSE, Interim Director for Patient Safety NHSI and Chair of the CSSPB, has asked for validation of this coding set.[3]

Establishing SOS outcomes will allow measurement of the impact of the NHS England strategies [4] (e.g. Sepsis CQUIN, Sepsis implementation guidance, Sepsis Action plans 1&2) in sepsis and infection. It would also provide a substrate for measurement of both the processes and impact of Public Health England strategies in raising sepsis awareness, infection control, antimicrobial stewardship and gram negative bacteraemia.

Furthermore, because SOS is the major cause of all cause deterioration that leads to emergency admission, it can be used to measure the impact of pan NHS implementation of National Early Warning Scores (NEWS). The dashboard work by the PSMU has revealed that pan regional implementation of NEWS is associated with an above expected reduction in SOS mortality in a 6 hospital region.

Validation of SOS would assist NHSE, NHSI and PSMU as we seek to develop an online tool for open access for all healthcare organisations, regions and affiliated bodies so they can develop their own SPC run charts and assess for the impact of local and regional improvement strategies around two key national work streams- deterioration and sepsis. It will be freely accessible via the PMSU website.

How to validate the SOS code bundle.

Though validation of a measurement tool may also be done by assessing consistency in reporting from multiple hospitals over time, this can only be part of the answer. It has to be supported by evidence of the accuracy of existing coding practice in interpreting clinical documentation of the objective clinical and laboratory data supporting a diagnosis of bacterial infection.

We have proposed a study to establish the validity of SOS by means of structured case note reviews to assess the level of agreement between the clinical reviewer's assessment of the evidence supporting presence or absence of bacterial infection, having reviewed all sources of information (vital signs charts, laboratory data, imaging reports, medication and fluid charts and medical notes, compared to the clinical coder's coding of diagnoses using the medical notes and discharge summaries. The clinical reviewer's assessment will be the gold standard. The structured questionnaire will constrain the temptation to second guess the judgment of the clinical team, by a rigorous systematic questionnaire asking objective questions reporting factual information from the available information. This should also minimise variation or bias in reporting between clinical reviewers.

The UK National Research Register was searched for similar studies that may be ongoing, but none were found. Through Google Scholar some precedents from the literature related to sepsis were found

where the use of routine hospital codes to track the epidemiology and outcomes of sepsis were validated by structured clinical notes review. Iwashyna et al[5] reported a study using a structured instrument to facilitate "charts" review (Charts having the equivalent meaning in the USA as medical records in the UK), to validate the use of the "Angus implementation" [6] of the International Consensus Conference definition of severe sepsis[7] to identify a hospital episode of sepsis by the use of sepsis specific ICD-9 CM codes. They concluded that using routine discharge codes was a "reasonable but imperfect" approach. We have studied the questionnaire developed by Iwashyna et al, adopted some relevant elements and adapted others in the process of developing our questionnaire.

There are though significant differences in context which means the Iwashyna study cannot be fully replicated. Firstly, the coding practices between the US and UK systems are different, in particularly in regarding the use of sepsis codes and the use of ICD-10 in the UK at present. Secondly the Angus criteria and Iwashyna's validation was focused on episodes of sepsis, and agreement that acute organ dysfunction was due to sepsis rather than another cause. The definitions of sepsis are still subject to change; since Iwashyna conducted the study, the 2001 consensus definition of sepsis (Sepsis-2) has been replaced by Sepsis-3[8], which abandons the Systemic Inflammatory Response Syndrome (SIRS) as too sensitive and non-specific, in favour of stricter demonstration of organ dysfunction. The changing definitions of sepsis make accurate epidemiological measurement over time difficult.

The focus of the SOS code bundle is on the broader concept of bacterial infection, and our study will focus on validating that episodes coded as bacterial infection are indeed either definite or probable cases of bacterial infection. Our expectation is that there is likely to be a much higher agreement between SOS classification of episodes as bacterial infections or otherwise based on coded diagnoses and clinical reviewers, than was the case with sepsis, as it should be much easier to determine if bacterial infection was proven or likely, than it was to attribute acute organ failure to sepsis or an alternative mechanism.

However, there is more to determine than just the level of agreement between clinical reviewers and coders. Validation has to assess whether any systematic errors are inherent to the process of reporting bacterial infection through routine hospital episode statistics. In our view the following systematic errors may affect the process:

a) Information bias; are clinical coders susceptible to under coding bacterial infection because they don't use sources of information that the clinical teams (and therefore the clinical reviewers) have

access to? They may be reliant on accurate clinical documentation which clearly label a definite diagnosis. For example, clinical coders would be unable to code "?cellulitis" or "possible cellulitis" but are able to code " probable cellulitis". Another example may be that in the absence of a stated diagnosis such as "chest infection", the clinical coder cannot document a diagnosis, but may resort to coding a ICD-10 coded symptom such as "fever" or "cough", which are not codes that are part of the SOS code bundle.

These type of errors will result in under coding true bacterial infection and therefore systematically reduce the sensitivity of the SOS code bundle.

b) Measurement bias; it is quite possible that in the absence of documentation of accurate statements summarising clinical assessment or laboratory evidence or indeed the absence of objective evidence of either a bacterial or viral infection, it is likely the viral infections may be coded as terms which may fall within the SOS code bundle. For example: "pneumonia" would be an SOS term, whereas "varicella pneumonia" would not be. The same would apply to overly generic terms such as "chest infection"

These types of errors would result in over coding SOS codes thereby reducing specificity.

6. PRELIMINARY STUDIES AND EXPERIENCE OF INVESTIGATORS

The investigators are familiar with this area of research. Dr Matt Inada-Kim, national sepsis advisor to the Department of Health and NHS Improvement, developed the SOS bundle of ICD-10 diagnosis codes signifying bacterial infection and published an epidemiological study using the classification system[1].

The investigator group at Portsmouth, including the chief investigator, have published a study comparing the performance of NEWS and qSOFA in infection versus no infection cohorts[2] using the SOS code bundle developed by Inada-Kim et al.

This preliminary work has already established the methods for correctly extracting and classifying hospital admissions into SOS and non-SOS cohorts.

7. AIMS AND OBJECTIVES

This study aims to assess whether the Suspicion of Sepsis (SOS) diagnosis code bundle is sufficiently accurate to identify patients with bacterial infections in hospital, by comparing it with the best available standard, case-note review. It also aims to identify and explore areas of discrepancy, between the two methods and within different hospitals.

7.1 Primary Objective

Blinded comparison of the accuracy of clinical coding in capturing episodes of bacterial infection in hospital admissions, assessed against a gold standard of structured clinical notes review.

7.2 Secondary Objectives

1. Assessment of true prevalence of bacterial infection and variation across sites.
2. Assessment of variation in practice of coding for bacterial infection between hospitals.
3. Assessment of the quality of clinical and laboratory evidence and the level of clinical conviction that bacterial infection had occurred.
4. Description of population differences between SOS and non-SOS cohorts summarising the data obtained from the structured case notes review.

7.3 Summary of Study Design

It will be a retrospective multi-centre observational study. At each site, two clinical reviewers who has lawful access to the medical records, will jointly conduct structured medical records review of randomly sampled hospital emergency admissions to agree whether or not a patient experienced a bacterial infection during hospitalisation.

An online questionnaire will lead the reviewers through a structured review of the records, which will comprise review of vital signs charts, laboratory tests, medication and fluid prescription charts, and clinical notation on paper or electronic patient record systems. The questions will extract objective data to aid assessment of the presence and quality of clinical and laboratory markers of infection, laboratory evidence of bacterial infection, level of clinical conviction as expressed by diagnostic statements and active treatment for bacterial infections. Finally, the clinical reviewers will be asked for a judgment whether the data supports assignment of the clinical episode to the bacterial infection

cohort. The reviewers' assignment to SOS or non-SOS cohort will be considered as the gold standard. Additional questions will determine whether patients had suffered any type of acute organ dysfunction, and whether in the reviewers' judgment the evidence suggested sepsis or alternative pathophysiological processes caused the organ dysfunction. The aim of these questions are to provide additional assurance the clinical context had been rigorously assessed.

The clinical reviewers will be blinded for the ICD-10 diagnosis codes assigned by clinical coders on the Patient Administration System (PAS) for the spell related to the hospital admission. The ICD-10 codes for each admission will be separately extracted and analysed to assign each spell to a bacterial infection (SOS) cohort or control (non-SOS) cohort based on the presence/absence of ICD-10 codes in the SOS bundle.

The study will not require active participation or consent of patients. The questionnaire will generate anonymised data at source from the medical record. Data from each completed questionnaire at the 5 sites will be uploaded to a single online database for analysis. The separately extracted ICD-10 codes and SOS cohort assignment based on those ICD-10 codes will be added to the database to allow comparison to the best available standard, clinical notes review. These separately sourced data will be joined in the database through the use of a pseudonymised study ID for each patient.

7.4 Outcome Measures

Primary outcome measure

The diagnostic accuracy between SOS coding and clinical notes review, where the clinical notes review will be the best available standard test.

Secondary outcomes measures

1. The prevalence of infection and its variation across the sites.
2. Measurement of variation between sites
    a. Variation in diagnostic accuracy, sensitivity and specificity.
    b. The distribution and variation in false positive rates and false negative rates across the sites as an assessment of systematic error in coding practices.
3. Quality of evidence on which diagnosis of bacterial information rests:

a. Proportions of episodes with evidence graded as “proven” i.e. there is definite evidence for bacterial infection, evidence consistent with bacterial infection but not diagnostic, episodes without any evidence or with evidence against bacterial infection.
b. Proportions of episodes with clinical team and reviewer conviction of the presence of bacterial infection graded as proven, probable or unlikely.

4. Summary statistics of population, treatment and outcome differences between SOS and non-SOS cohorts with means, medians and proportions reporting 95% confidence intervals and significant tests with a type I error level of 0.05 for variables described in section 6.2.3a to 3d. These may include but are not limited to:
    a. Differences in characteristics such as age, gender, presenting symptoms, comorbidities, risks for infection, physiological markers of deterioration (NEWS)
    b. Differences in management such as antibiotic prescribing, requesting and reporting of microbial diagnostic tests, Hospital@Night or Outreach team review, Critical Care team review, Intensive Care admission, or End of Life care decisions.
    c. Prevalence, type, severity (number of organs) and pathophysiological process attribution of acute organ dysfunctions occurring in the SOS cohorts non-SOS cohort as a means to provide important SOS population description.
    d. Differences in outcomes such as death and cardiac arrest.

8. STUDY PARTICIPANTS

8.1 Study Setting

The study will be conducted at five large NHS Trusts in England, distributed throughout the country (Portsmouth Hospitals NHS Trust, Hampshire Hospitals NHS Foundation Trust, Cambridge University Hospitals NHS Trust, Liverpool and Broadgreen University Hospitals NHS Trust, and South Tees Hospitals NHS Foundation Trust). These Trusts are representative of the variation in hospitals serving city, town and rural populations in the north and south of England, and include both tertiary/teaching hospitals and district general hospitals.

8.2 Overall Description of Study Participants

For this retrospective study reviewing medical records there will be no active study of patients directly as participants. Instead the focus is on the quality of evidence recorded in the medical records of randomly sampled hospital stays of patients. The random sample strategy is described in Section 9.

8.3 Eligibility Criteria

Inclusion Criteria

An admission must meet ALL of the following criteria to be considered eligible for the study:

- Adult, aged ≥16
- elective or emergency admission to one of the five participating hospitals between April 1st, 2017 and March 31st, 2018
- discharged alive or dead at the time of data extraction, with complete consultant episodes recorded on PAS.

Exclusion Criteria

- Length of stay of zero days (i.e. no overnight stay).

9. SAMPLING

Sampling of admissions for the year 2017/18 for case note review will be random, stratified by month to avoid seasonal bias. The list of admissions for each month will be produced and processed separately to facilitate monitoring and validation, maintain correct stratification and reduce the chances of mistakes.

The process of generating the random samples at each site is described in Section 10.1 according to work instruction WI01.

Sample size calculation for the primary outcome of the level of agreement between SOS coding and clinical notes review (accuracy) is based on the required precision of 2.5% on an estimated level of agreement of 90%.

The total number of notes needed to provide precision of 2.5% with an expected 90% proportion of agreement is 554. This number rises to 784 for an 85% accuracy. Please review Appendix B for the detailed sample size modelling based on a range of SOS prevalence estimates and required diagnostic accuracy, sensitivities and specificities for the whole study sample and site samples.

It has been agreed to review 144 notes per site (a total of 720). This number will allow for:

- 5% precision estimates on the overall accuracy at each site (for an expected 90% agreement). Arguably validity of using SOS diagnosis codes if diagnostic accuracy is significantly below 90% is questionable.
- precise estimates of sensitivity and specificity for the whole sample (better than 5% precision at levels of 85% and 75% respectively). It is expected that sensitivity will be higher than specificity, as it is likely that clinicians will err on the side of caution to diagnose and treat any episodes of febrile illness as a bacterial infection in the absence of evidence to the contrary eg objective evidence of a viral infection.
- Assessment of the medical records of 12 admissions in each of the 12 months of the study period.

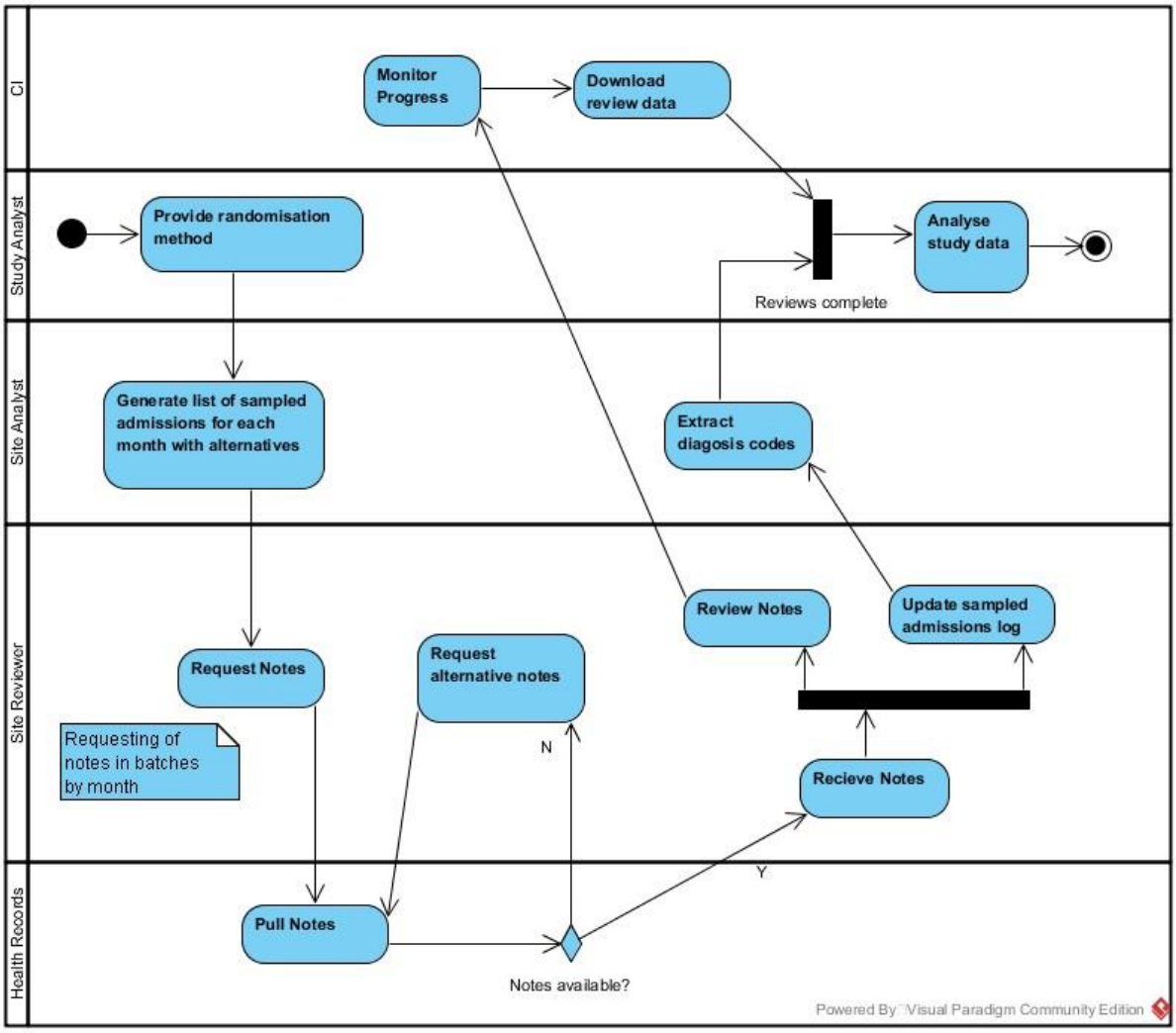
We have assumed prevalence of bacterial infections amongst hospital inpatients, either as the cause of admission, or acquired in hospital, fall somewhere between 25-38%, more likely greater than 30% o as the PSMU has been reporting increasing prevalence of SOS diagnoses.[9] We modelled the impact of different prevalence levels to understand how it could affect the sample size required. We have

elected to sample significantly more than the minimum number of notes to measure whether there is at least 90% agreement with a 5% precision estimate at the site level.

It has been further agreed that review of individual medical records would be pragmatically curtailed at 14 days following admission, for the few admissions that are in hospital for a prolonged period, Commensurate with this decision, the diagnosis codes will not be extracted for Finished Consultant Episodes (FCEs) starting after 14 days in hospital. The number or FCEs affected or excluded are unlikely to affect answering the primary research question and objective.

10. STUDY PROCEDURES

No patient consent will be required prior to data collection from selected case notes. This aspect is further expanded on in Section 13 (Ethics). The overall process controlling the data acquisition is set out in the diagram below and further explained in Sections 10.1 to 10.3.



10.1 Selection of notes

The site medical records manager will be contacted in advance by the site's research department and briefed on the study requirements so that appropriate arrangement can be made, e.g. handle it similarly to a standard local clinical audit.

Each site analyst will generate the whole sample of 144 case notes immediately after site initiation, divided into 12 monthly sample lists at site initiation according to work instruction WI01 for the study. The lists will be configured as separate spreadsheets where each list is in the form of a sample log for

the month's admissions which the reviewers will update as they progress through the reviews. Each list will indicate two alternative case notes that could be used should the notes for some of the admissions not be available.

All 12 lists will be given to the Principle Investigator (PI), who is also the lead clinical reviewer who will manage it in accordance with WI02 for the study.

WI01 – Creating the lists of sampled admissions

Undertaken by site analyst using corporate reporting tools and a VBA program for Microsoft Access written by the study analyst to ensure all sites follow exactly the same methodology.

1. Query PAS/EPR for discharged and coded admissions for patients over 16 years of age on their admission date where the admission date is between 1st April 2017 and 31st March 2018 and the admission date and discharge date are different (i.e. there was an overnight stay). Load admissions into Access table "SOS_admissions".
2. Run form "Main" in the Access database.
3. Click the "Validate eligible admissions" button. Any invalid admissions will be displayed. Resolve, repeat and recheck as required before continuing.
4. Click the "Make table of index admissions" button. This creates the table "SOS_Index_admissions" where all admissions relate to different individuals. (Contains only the first admission in the year for each.) ***This ensures no individual patient is sampled twice.***
5. Click the "Create Lists" button. The Access VBA program will create 12 tables from SOS_admissions with names "SOS_monthname". A column "AdmNo" is added to each table and autonumbers the admissions for each month ordered by date and time. A message box will display after each month, showing the number of admissions for each month.
5. The VBA program will also use a random number generator to randomly sample 36 numbers from each monthly table, producing another 12 tables with the names "SOS_monthname_sample". ***These tables generate the study id for each admission*** in the form X/MM/NNN where X is the pseudonymised site id, MM is the month number, NNN is the row number provide by the new autonumber field.
6. Click the "Export lists" button. This exports the monthly sample lists to Excel spreadsheets in the same folder as the database is in.
7. Check that the first 12 admissions in each list are random over the month by admission datetime.

7. Export monthly sample list tables to Excel spreadsheets, lock the populated fields so that the data cannot be inadvertently, set a password for the Excel file and send to site reviewers by secure email.

10.2 Data extraction from clinical notes

The survey will be conducted using a web based survey management application further described in Section 11. The survey content is described in Appendix 1.

WI02 – Managing the survey completion process

For participating sites with fully electronic medical records, a simpler notes review process will ensue, as clinical reviewers will simply be able to progress with the first 12 selected episodes for each month of the year until all 144 selected record reviews have been completed.

For sites where paper medical records are still being used, clinical reviewers will manage the clinical notes review process as follows:

1. It is recommended that the reviewers will enlist the assistance of a medical secretary or other trusted administrative staff (working for the PI or department) as they are usually very familiar with the process of requesting notes from the medical records department, and can ensure that there is no delay in requesting the next batch of notes, or returning the completed batch.
2. The records will be requested in batches by month (e.g. April 2017, then May 2017, etc.) and returned according to local procedures. This will avoid delays e.g. by requesting on month batch in advance as a buffer held by the administrative staff, and will prevent the wrong set of notes inadvertently being selected.
3. The notes for each month will be requested as a batch of 12 notes, using the 36 selected case notes number sequentially i.e. the first 12 numbers on the list. Where a particular set of notes are not immediately available (e.g. not in the library), the reviewers/delegated administrative staff will request an alternative set of notes using the next number (i.e. the 13th and so forth on the sample admission list) from the same month.
4. Administrative staff will check that no episode coding is being displayed in the notes. If it is, it will be temporarily masked by an A4 sheet of paper.

5. The clinical reviewer will, upon starting the online survey for a selected episode
    a. select an online survey from 144 emailed links to uniquely numbered surveys, noting the unique survey identifier.
    b. Select a set of notes from the current batch, checking that the case note number and date/time of admission is correct. If so, update the Excel spreadsheet log for the month by adding 1) the unique survey number 2) the name of the reviewer and 3) patient initials from the notes in separate columns on the spreadsheet to the row of the corresponding case note number. If not correct, log this on the Excel spreadsheet, and request the next set of case notes on the list.
    c. Check the admission and discharge dates. If a patient remained in hospital continuously for more than 14 days, curtail charts and notes reviews at 14 days. Check the survey item indicating a curtailed review.
    d. Complete the online survey, starting by entering 1) the patient initials and 2) the study ID in the form X/MM/NNN corresponding to the case note under review. At any point in the review the reviewer can save a draft version, for example when interrupted. Upon completion, a summary of the survey with answers to each question will be published.
6. The case notes will then be passed to the second clinical reviewer, who will review the notes and survey summary, while the primary reviewer of the notes summarise the particular admission, and the evidence found supporting particular answers. The focus of the second reviewer will be on inconsistencies or omissions in the evidence, and the validity of conclusions drawn. There should be a discussion to achieve agreement on points of contention. If agreement cannot be achieved, the case should be passed to a third (senior) nominated clinical reviewer for adjudication. Once the process is completed, the final version can be submitted.

10.3 Data extraction from PAS

1. A copy of the completed Excel sample log, with all data fields locked to prevent inadvertent altering of data, will be returned to the site analyst who will use them to query the site's PAS for the consultant episode diagnoses for the sampled admissions whose notes have been reviewed.

2. At this point the site analyst will double check that the documented initials of the patient correspond to the name of the patient whose consultant episode details are being extracted, and proceed to prepare a data file containing:

    a. The study ID number described in Section 10.1.6

    b. Patient initials

    c. Patient gender

    d. Patient age

    e. Method of admission

    f. Month of admission

    g. Weekday of admission (1- Monday, 2- Tuesday etc)

    h. All finished consultant episodes (FCEs) with all diagnoses codes and each diagnosis number) and Admission day number of start of consultant episode.

    i. An indicator if any FCEs extracted *bridged 14 days* 0=no; 1=yes (NB Note the difference: Indicator should not be used to indicate FCEs starting after 14 days that had been excluded from the extract).

3. The Access file will be passed on to the study information analyst, Dr Paul Meredith at Portsmouth Hospitals NHS Trust to process as follows to further de-identify the final study database.

    a. Once the clinical coding extract and case notes review data has been joined and checked to ensure that the patient initials and study ID agrees for both sources of data, the patient initials will be deleted.

    b. Summary descriptors for age such as average and median age, and age range for both SOS and non-SOS cohorts will be calculated. Patients will be allocated to 5-year age bands and patient age will be deleted from the final study database.

11. DATA HANDLING AND RECORD KEEPING

11.1 Data Collection Forms and databases

The case note review tool has been developed by the Chief Investigator. The design process went through the following stages:

- Review of published surveys with the same aims i.e. validating the use of clinical coding of hospital treatment episodes for the purpose of tracking the prevalence and outcomes of bacterial infections in the United Kingdom
- We only found 1 validated structured survey by Iwashyna et al, which had a similar, not identical aim of validating clinical coding of hospital episodes to track the prevalence and outcomes of sepsis in the United States of America.
- The survey designed by Iwashyna was examined and adapted to our study aims and objectives, and health care context in England.
- It was then subjected to peer review by the study investigators, and finally external review by the Sepsis Advisory Panel for NHS England.
- The final version of the questionnaire is attached as Appendix A to the protocol

The case note review tool will be delivered as an online questionnaire using the on-line Smart Survey tool, https://www.smartsurvey.co.uk/. The survey questions detailed in Appendix A will be converted to online forms by the CI. The tool data is stored and backed up on UK/EU-based servers and is hosted by an ISO 27001 certified company. The application is complaint with GDPR standards.

The data entered will be pseudonymised using study identifiers as outlined in Section 10.1 (WP01 description) and will be effectively anonymous according to the ICO code of practice.

The PAS coding records collection database will be an Access database designed by the study analyst. It will comply with the NHS data dictionary in terms of the data items it collects. It will be tested using data extracted from the PHT PAS prior to study use.

The study database for data integration will be designed and developed in Access by the study analyst. Its development will follow testing of the case note review tool since its design will be based on the output from that tool.

11.2 Data Management

A study database will be created by the study analyst to integrate the findings from the structured case notes reviews and corresponding pseudonymised ICD-10 diagnoses codes ordered by consultant episode as recorded on each site's PAS data, and prepare it for analysis.

12. DATA ANALYSIS

12.1 Description of Analysis Populations

The only eligibility criteria for inclusion in analysis is that the patient should have been at least 16 years old at the time of admission between the 1st of April 2017 and the 31st of March 2018. For inclusion in the SOS cohort, the patient should have had at least one SOS code in any position in any of the FCEs that constitute the admission Any episode that does not satisfy the latter criterion will fall within the non-SOS cohort. The baseline characteristics of the analysis populations will be compared. FCE diagnoses data will be explored to establish if limiting the SOS cohort to codes in higher positions (for example 1st and 2nd position) in each FCE, or restricting the analysis to the admission or discharge FCE makes a notable difference to the level of agreement or to cohort allocation.

12.2 Analysis of Endpoints

The primary measure is diagnostic accuracy of SOS categorisation derived from diagnostic codes as determined by the clinical conviction of bacterial infection evidenced by the case note reviews. Accuracy is defined as the proportion of agreement between SOS categorisation and case note review. This will be reported as a whole and by site with 95% confidence intervals. The sensitivity (proportion of positive cases identified) and specificity (proportion of negative cases identified) will be reported similarly.

Analysis of secondary outcome measures

All proportions will be reported with 95% confidence intervals.

1. Prevalence of infection as determined by the case notes reviews as a proportion of the sampled admissions will be reported with confidence intervals for the whole data set and by site. Variation between sites will be tested with a chi-squared test of goodness of fit of the prevalences by site. When the null hypothesis is rejected, significant pair-wise differences will be identified by Marascuilo's procedure.[10]
2. Site variation analysis
    a. Variations in diagnostic accuracy, sensitivity and specificity by site will similarly be reported as proportions with confidence intervals, tested for significant variation with a chi-squared test and significant pair-wise differences identified by Marascuilo's procedure.

b. the false positive and false negative rates and their confidence intervals will be calculated for each site. Again, significant variation with will tested with a chi-squared test and significant pair-wise differences identified by Marascuilo's procedure. The systematic bias demonstrated by false positives and false negatives will be reported as an odds ratio with confidence intervals for each site.

3. Quality of evidence for bacterial infection

    a. The proportions of patients with categorised clinical and laboratory evidence for bacterial infection will reported for each site with confidence intervals. A chi-squared tested of independence will be used to test for a null association between the three proportions of evidence and the sites.

    b. Similarly compare the three categorisations of admissions by clinical conviction of bacterial infection by site.

4. Summary statistics of population, treatment and outcome differences between SOS and non-SOS cohorts with means, medians and proportions reporting 95% confidence intervals and significant tests with a type I error level of 0.05 for variables described in section 6.2.3a to 3d

12.3 Procedure for Dealing with Missing, Unused and Spurious Data

Missing data will be minimised by the use of online structured survey tools. The survey cannot be completed with missing data. It will be assumed that if clinical reviewers find that particular tests were not done by clinical teams during the diagnostic process, that it was appropriate and not deemed necessary to the management of the patient, rather than that the data is missing. Therefore, no imputation of missing data will be performed.

13. ETHICS

The main ethical consideration for this study is whether individual patient consent is required for each admission spell that will be selected. Advice we have taken is that this is not required for the following reasons:

- All data in the medical records accessed for this study has been collected for routine care purposes that have completed and is held by each study hospital as medical records covered by existing legal arrangements.

- These data will be accessed by clinical staff who have existing permission to access the medical records of any patients admitted to the hospital in order to fulfil duties such as the provision of clinical care or clinical audit.
- The data will be anonymised in two stages so that the study investigators will not be able to identify any individual patient from any individual completed survey from or the study database.
    - The first stage occurs on site, so that patient identifiers such as name, date of birth or hospital number is replaced by a generated study ID, age and patient initials
    - Only the patient initials are entered into the online survey, and only age and patient initials are extracted from the local PAS and forwarded to the study analyst, Dr Paul Meredith, by password secured email.
    - The second stage of removing patient identifiers occurs when the data from the survey and the PAS extract containing the coded diagnoses has been joined. The patient initials is used as a check that the correct medical record had been surveyed. Once this has occurred, the patient initials is removed from the study database. Age is replaced by 5-year age band in the study database once the population summary statistics have been calculated.

We cannot foresee any adverse effects on individual patients for being included in the study. Vulnerable patients or groups will have an equal chance of being selected for the study as their representation in the adult hospital population.

A retrospective case notes review could discover evidence of clinical negligence. However, this structured survey is set up to document what had happened, not what should have happened, therefore it is not a suitable tool for achieving judgment on such cases. In addition, NHS Trusts now have robust tools (Mortality Review Panels) in place to examine extent to which of each death in hospital had been avoidable. In the highly unlikely event clinical negligence is discovered, it will be the duty of the Principle Investigator to record it as a potential adverse incident on the individual Trust's reporting tool for further investigation and grading by the clinical governance team.

13.1 Participant Confidentiality

The study will comply with GDPR/Data Protection Act 2018 which requires data to be anonymised as

soon as it is practical to do so. Clinically sensitive data will be anonymised at source (by the clinical reviewers) and a quality controlled process of removing any direct patient and episode identifiers will occur on each study site. The investigators will ensure that the participants' anonymity is maintained by following the protocol which sets out the process of de-identifying the study database. Participants will be identified by initials and study ID number on any electronic forms or databases. Participant initials will be removed as soon as the data joining step has been completed. All paper or electronic documents related to the study will be stored securely and only accessible by study staff and authorised personnel.

13.2 Declaration of Helsinki

The study will be conducted according to the principles of the Helsinki declaration. We believe the trial design honours the requirement to maintain privacy and confidentiality of individual patients. Patient consent or dissent is impractical to obtain, but as patient identifiable date will not be used, the data collected concerns their health status at a particular time in the past, and is derived data which has no sensitive content, we believe this can be justified by absence of potential harm to any patient.

13.3 ICH Guidelines for Good Clinical Practice

Portsmouth Hospitals NHS Trust R&I department will require that all study investigators will have up to date GCP certification and will monitor the study for adherence to GCP.

14. PATIENT PUBLIC INVOLVEMENT (PPI)

The study was commissioned by NHS Improvement to validate a proposed measurement standard. No direct patient or public involvement in the design of the study or survey was required by the funder, though it is possible that lay members of the public or patients inform NHS Improvement as part of the decision making process. NHS Improvement will lead on deciding the extent of communication to the public about the study results, the resources needed to do so and channels through which the information would be disseminated, other than the report and peer reviewed journal publication.

15. FINANCING AND INSURANCE

The study is funded by NHS Improvement for a total of £40,000. Each participating site is allocated £5000 to cover medical records department expenses for retrieval (£10/per set of notes), site analyst

time to generate the random sample and extract date, and clinical reviewer time (estimated at 36 hours per clinical reviewer at each site).

£15000 in funding is assigned to Portsmouth R&I department to deliver the study, consisting of protocol development, IRAS and CAG approval if needed, administration, contracting with each participating site, creating study databases, conduct analysis and produce final report.

16. TIMETABLE AND ORGANISATIONAL CHART

Milestones

- Finalisation of Protocol and survey
- Obtaining HRA and Ethics approval
- Study set up in each participating site
- Recruitment complete in all sites
- Manuscript drafted and submits final report to NHS improvement

Gantt chart indicating the timeline for work packages to meet study milestones

	Work packages/Actions	Sept	Oct	Nov	Dec	Jan	Feb	March
1	Protocol and survey development							
2	HRA and Ethics							
3	Study set up in each participating site							
4	Recruitment/data collection							
5	Manuscript and report preparation							
6	Monthly report update to funder	31-Sept-2018	28-Oct-2018	26-Nov-2018	30-Dec-2018	28-Jan-2018	28-Feb-2019	31-March 2019

17. RESOURCES, EQUIPMENT AND PHYSICAL FACILITIES

Each site would require the availability of:

- An information analyst with access to the PAS system, familiarity with the PAS data dictionary, competency to extract data to and manipulate date in Microsoft Access database tables and Excel spreadsheets.
- A consultant and a clinical colleague (senior or junior qualified medical practitioner,) with access to an office where medical notes can be securely kept, and equipped with at least two desktop or laptop computers with access to the internet.

- Medical records requesting facility to retrieve and deliver batches of notes to the secure office.

Estimates of time commitment at each site to complete the notes review on time.

- Information analyst.
  - 0.5 day to prepare the random, seasonally adjusted sample, add study IDs, and create 12 monthly lists in Excel.
  - 0.5 day to extract PAS data for the final site sample (for which notes review has been completed) to an Access database, and prepare the database to send to the study information analyst.
- Medical records clerk
  - Time required to retrieve and deliver 144 sets of notes in batches of 12 over a period of 3 to 4 months.
- Lead consultant clinical reviewer and second clinical colleague each:
  - Time required to complete survey for 72 notes (primary review role) – approximately 20 minutes per set of notes
  - Time required to participate in discussion with second reviewer while second reviewer reviews above 72 sets of notes and answers to check validity of answers – 10 minutes per set of notes
  - Time required to be the second reviewer discussing and checking the reviews the other primary reviewer did – 1 0 minutes per set of notes.
  - Estimated time requirements will be 4 PAs of SPA time per month for 3 months between September and December 2018 for each reviewer.

18. DISSEMINATION AND OUTCOME

A Memorandum of understanding has been agreed between NHS Improvement and Portsmouth Hospitals Research & Innovation department stipulating that:

- The Investigator team should produce a report summarising the results of the study for internal dissemination to chosen agencies of NHS England and the Department of Health. The original agreed date was the 31st of January, 2019, but either party may ask for an extension. Delays in obtaining external reviews and responding to the feedback, has pushed out the study start date by two months. However, as far as possible we will endeavour to make up for lost time to conclude the study as quickly as feasible.
- The investigator team will simultaneously prepare a paper for publication in a peer reviewed journal appropriate to the subject matter, and reflecting the relevance of the findings to a clinical, academic and health systems research and management audience interested in sepsis. The aim will be to minimise delay between producing the internal report and external publication of the findings.
- Reporting will follow STROBE guidelines[11] for reporting observational studies.

The nature of the study and resulting report is such that if it supports the adoption of SOS by validating it's use, it would be impactful, and likely to affect policy decisions in the UK countries and internationally how to measure incidence of infections, and changes in outcomes form from sepsis in response to quality improvement initiatives.

19. REFERENCES

1. Inada-Kim, M., et al., *Defining and measuring suspicion of sepsis: an analysis of routine data.* BMJ Open, 2017. **7**(6).
2. Redfern, O.C., et al., *A Comparison of the Quick Sequential (Sepsis-Related) Organ Failure Assessment Score and the National Early Warning Score in Non-ICU Patients With/Without Infection.* Critical Care Medicine, 2018. **Online First**.
3. Inada-Kim, M. 2018.
4. Ingham Clark, C., *Cross-system sepsis action plan 2017*, N. England, Editor. 2017, NHS England.
5. Iwashyna, T.J., et al., *Identifying Patients with Severe Sepsis Using Administrative Claims: Patient-Level Validation of the Angus Implementation of the International Consensus Conference Definition of Severe Sepsis.* Medical care, 2014. **52**(6): p. e39-e43.
6. Angus, D.C., et al., *Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care.* Critical Care Medicine, 2001. **29**(7): p. 1303-1310.
7. Levy, M.M., et al., *2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference.* Intensive Care Med, 2003. **29**(4): p. 530-8.
8. Singer, M., et al., *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3).* JAMA, 2016. **315**.
9. Imperial College Health Partners. *Suspicion of Sepsis (SOS) Insights Dashboard*. 2018 22/09/2018]; Available from: https://www.sos-insights.co.uk/.
10. Wagh, S.T. and N.A. Razvi, *Marascuilo method of multiple comparisons (an analytical study of caesarean section delivery).* Int J Contemp Med Res, 2016. **3**(4): p. 1137-40.
11. von Elm, E., et al., *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies.* International Journal of Surgery, 2014. **12**(12): p. 1495-1499.

20. APPENDIX A STRUCTURED SURVEY TOOL

Structured Survey tool uploaded as a separate document

21. APPENDIX B SAMPLE SIZE CALCULATIONS

Whole study sample = 720 Site samples = 144

Precision of estimates for different thresholds of diagnostic accuracy (level of agreement)

Accuracy sample size calculations for proportion p with precision r (same for all prevalences)											
		r									
		1.0%	2.0%	2.5%	3.0%	4.0%	5.0%	6.0%	7.0%	8.0%	9.0%	10.0%
p	0.75	7203	1801	1153	801	451	289	201	147	113	89	73
	0.8	6147	1537	984	683	385	246	171	126	97	76	62
	0.85	4899	1225	784	545	307	196	137	100	77	61	49
	0.9	3458	865	554	385	217	139	97	71	55	43	35
	0.95	1825	457	292	203	115	73	51	38	29	23	19

Precision of estimates for Sensitivity and specificity at SOS Prevalence 32.5%

Sensitivity sample size calculations for proportion p with precision r and prevalence 32.5%											
		r									
		1.0%	2.0%	2.5%	3.0%	4.0%	5.0%	6.0%	7.0%	8.0%	9.0%	10.0%
p	0.75	22164	5541	3547	2463	1386	887	616	453	347	274	222
	0.8	18913	4729	3026	2102	1183	757	526	386	296	234	190
	0.85	15071	3768	2412	1675	942	603	419	308	236	187	151
	0.9	10639	2660	1703	1183	665	426	296	218	167	132	107
	0.95	5615	1404	899	624	351	225	156	115	88	70	57
Specificity sample size calculations for proportion p with precision r and prevalence 32.5%											
		r									
		1%	2%	3%	3%	4%	5%	6%	7%	8%	9%	10%
	0.5	14229	3558	2277	1581	890	570	396	291	223	176	143
	0.55	14086	3522	2254	1566	881	564	392	288	221	174	141
	0.6	13660	3415	2186	1518	854	547	380	279	214	169	137
	0.65	12948	3237	2072	1439	810	518	360	265	203	160	130
p	0.75	10672	2668	1708	1186	667	427	297	218	167	132	107
	0.8	9107	2277	1457	1012	570	365	253	186	143	113	92
	0.85	7257	1815	1162	807	454	291	202	149	114	90	73
	0.9	5123	1281	820	570	321	205	143	105	81	64	52
	0.95	2704	676	433	301	169	109	76	56	43	34	28

Precision of estimates for Sensitivity and specificity at SOS Prevalence 38%

Sensitivity sample size calculations for proportion p with precision r and prevalence 38%											
		r									
		1.0%	2.0%	2.5%	3.0%	4.0%	5.0%	6.0%	7.0%	8.0%	9.0%	10.0%
p	0.75	18956	4739	3033	2107	1185	759	527	387	297	235	190
	0.8	16176	4044	2589	1798	1011	648	450	331	253	200	162
	0.85	12890	3223	2063	1433	806	516	359	264	202	160	129
	0.9	9099	2275	1456	1011	569	364	253	186	143	113	91
	0.95	4802	1201	769	534	301	193	134	99	76	60	49
Specificity sample size calculations for proportion p with precision r and prevalence 38%											
		r									
		1%	2%	3%	3%	4%	5%	6%	7%	8%	9%	10%
	0.5	15491	3873	2479	1722	969	620	431	317	243	192	155
	0.55	15336	3834	2454	1704	959	614	426	313	240	190	154
	0.6	14871	3718	2380	1653	930	595	414	304	233	184	149
	0.65	14097	3525	2256	1567	882	564	392	288	221	175	141
p	0.75	13012	3253	2082	1446	814	521	362	266	204	161	131
	0.8	11618	2905	1859	1291	727	465	323	238	182	144	117
	0.85	9914	2479	1587	1102	620	397	276	203	155	123	100
	0.9	7901	1976	1265	878	494	317	220	162	124	98	80
	0.95	5577	1395	893	620	349	224	155	114	88	69	56