TITLE PAGE

Division: Worldwide Development
Information Type: Worldwide Epidemiology Study Protocol

Title:	PRJ2282 / 201491: CHESS: CPRD-COPD Hawthorne Effect Study in Salford: A UK cohort study to characterise patients enrolled in the Salford Lung Study and to evaluate a potential Hawthorne effect

Compound Number:	GW685698 / GW642444
Development Phase	IV
Effective Date:	23 October 2015
Subject:	COPD in the UK general population and comparison with the Salford Lung Study
Author(s):	**University of Manchester:** Sperrin, Matthew; Van Staa, Tjeerd; **GlaxoSmithKline:** Davis, Kourtney; Pimenta, Jeanne; **Clinical Practice Research Datalink:** Williams, Rachael.

TABLE OF CONTENTS

PAGE

1. LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
CHESS	CPRD-COPD Hawthorne Effect Study in Salford
CPRD	Clinical Practice Research Datalink
COPD	Chronic Obstructive Pulmonary Disease
EHR	Electronic Health Record
EMR	Electronic Medical Record
FEV1	Forced expiratory volume in 1 second
FF	Fluticasone furoate
FVC	Forced vital capacity
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GPRD	General Practice Research Database
GP	General Practitioner
GSK	GlaxoSmithKline
HCU	Health Care Utilisation
HES	Hospital Episode Statistics
ICS	Inhaled Corticosteroid
ISAC	CPRD Independent Scientific Advisory Committee
LABA	Long Acting Bronchodilator
LAMA	Long Acting Muscarinic Antagonist
MHRA	Medicines and Healthcare Products Regulatory Agency
MPR	Medication Possession Ratio
MRC	Medical Research Council
NICE	National Institute for Health and Care Excellence
NWeH	North West eHealth
ONS	Office for National Statistics
PDC	Percent Days Covered
SAS	Statistical Analysis System
SES	Socio-Economic Status
SIR	Salford Integrated Record
SLS	Salford Lung Study
SOC	Standard of Care

SOP	Standard Operating Procedure
UoM	University of Manchester
VI	Vilanterol

Trademark Information

Trademarks of the GlaxoSmithKline group of companies
RELVAR

Trademarks not owned by the GlaxoSmithKline group of companies
SAS

2. RESPONSIBLE PARTIES

The study will be led by University of Manchester (UoM) who will collaborate with North West eHealth (NWeH), Clinical Practice Research Datalink (CPRD) and GSK.

Study Task	**Responsible group(s)**
Protocol finalization	UoM
Develop specifications for data sets	UoM, GSK, CPRD, NWeH
Data extraction for SLS Standard of Care arm	NWeH
Creation of CPRD cohort dataset	UoM, CPRD
Analyses (using SAS) comparing SLS and CPRD	UoM
QC/QA of analysis	GSK (Observational Data Analytics)
Final study report	Drafting: UoM Review, comment, and edits: Scientific Committee Reporting to regulatory agency and web-based register: GSK
Manuscripts	Drafting: medical writer Review and edits: study team, scientific committee members

A Scientific Committee (see section below) will be assembled and will be required to review and input into study design and major study documents (final study protocol, final research analysis plan, final study report, peer-review publications).

SPONSOR SIGNATORY:

Jeanne M. Pimenta Primary Author / Project officer	**Date**
Kourtney J. Davis Therapy Area Leader (Respiratory), Worldwide Epidemiology)	**Date**
Andrew W. Roddam Vice President, Worldwide Epidemiology	**Date**

SPONSOR INFORMATION PAGE

WWEpi Project Identifier:

PRJ2282

eTrack Project Identifier:

201491

Sponsor Legal Registered Address:

GlaxoSmithKline Research & Development Limited
980 Great West Road
Brentford
Middlesex, TW8 9GS, UK

Sponsor Contact Address

GlaxoSmithKline Research & Development Limited
Ironbridge Road
Stockley Park West, Uxbridge,
Middlesex, UB11 1BU, UK
Telephone: +44 (0)208 990 9000

INVESTIGATOR PROTOCOL AGREEMENT PAGE

- I confirm agreement to conduct the study in compliance with the protocol.
- I acknowledge that I am responsible for overall study conduct. I agree to personally conduct or supervise the described clinical study.
- I agree to ensure that all associates, colleagues and employees assisting in the conduct of the study are informed about their obligations. Mechanisms are in place to ensure that site staff receives the appropriate information throughout the study.

Investigator Name: **Matthew Sperrin**

28/10/15

Investigator Signature | Date

STUDY SCIENTIFIC COMMITTEE

External Members:

Darren Ashcroft

Professor of Pharmacoepidemiology, School of Pharmacy and Pharmaceutical Sciences, University of Manchester

Iain Buchan

Clinical Professor in Public Health Informatics, Health eResearch Centre, Farr Institute for Health Informatics Research, University of Manchester

Tjeerd van Staa

Professor of Health eResearch, Health eResearch Centre, Farr Institute for Health Informatics Research, University of Manchester

John New

Chief Clinical Officer, North West eHealth, Salford Royal Hospital

Liam Smeeth

Professor of Clinical Epidemiology, London School of Hygiene and Tropical Medicine

Peter Lange

Professor, Consultant in Respiratory Medicine, Department of Public Health, Section of Social Medicine, University of Copenhagen

Rachael Williams

Research Statistician, Clinical Practice Research Datalink

Tim Williams

Head of Research, Clinical Practice Research Datalink

GSK Members:

Kourtney Davis

Senior Director, Worldwide Epidemiology, Respiratory Therapeutic Area Lead, Worldwide Epidemiology GlaxoSmithKline

David Leather

Global Medical Advisory Lead, Respiratory Franchise, GlaxoSmithKline

Sally Lettis

Director, Clinical Statistics, GlaxoSmithKline

3. ABSTRACT

Background: The Salford Lung Study (SLS) is a unique Phase IIIB pragmatic trial evaluating the effectiveness of a novel medicine – RELVAR (fluticasone furoate [FF], GW685698) and vilanterol [VI] GW642444) – compared with standard of care (SOC) among patients with Chronic Obstructive Pulmonary Disease (COPD). The trial is taking place in Salford, England. While the pragmatic nature of the trial is designed to test effectiveness in routine care, there are at least two possible concerns: 1) Salford may not be representative of the wider population in which the medicine may be used, and 2) there may be differences in local practice or changes to local practice caused by the study (the Hawthorne effect), which may artificially inflate the benefits of both RELVAR and SOC.

Objectives: The aim of the study is to evaluate the representativeness of Salford, and the potential Hawthorne effect, to place the SLS in wider context. The objectives are:

Co-primary objectives:

- **PO1**: To characterize the patients enrolled in the Standard of Care (SOC) arm of SLS COPD compared with the UK population of COPD patients (using the Clinical Practice Research Datalink (CPRD)), including distribution of SES/deprivation level, to evaluate the extent to which the SLS participants are representative of the UK patient population targeted for RELVAR. The comparator set will be specified on two bases: firstly, overall, and secondly, the subset fulfilling the protocol inclusion/exclusion criteria.

- **PO2**: To compare the rate of COPD exacerbation over the 12 months in Standard of Care arm of the SLS compared with the Standard of Care (SOC) recorded in the CPRD, in order to detect a potential Hawthorne effect.

- **PO3**: To compare the rate of serious pneumonia (defined by hospitalisation) over the 12 months in Standard of Care arm of the SLS compared with the Standard of Care recorded in the CPRD.

Secondary objectives:

- **SO1**: To make comparisons between the SLS SOC and the CPRD cohort on the following health care utilisation (HCU) endpoints: GP visits, hospital admissions, mortality and adherence.

- **SO2**: To evaluate other definitions of COPD exacerbations in SOC from CPRD.
.

- **SO3**: Self-controlled comparison of COPD and HCU endpoints in Salford before and after SLS commenced, using data from the SLS.

Study Design: Observational COPD cohort study, comparing Salford, UK (data source: SLS) with rest of UK (data source: CPRD) over a 12 month period.

Population: The setting is Salford, UK compared with rest of UK. The study population is, in Salford, participants recruited to the SLS and randomised to the SOC arm. In the rest of the UK, the comparison group is persons with COPD recorded in the CPRD who meet the eligibility criteria of the SLS.

Study Size: The target sample size for the SLS is 2,800. Study size in the CPRD will be based on the prevalence of COPD diagnosis codes; this will be a minimum 2,800 to match but is expected to be considerably larger (x5-10).

Analysis: Data analysis will use descriptive statistics for PO1. PO2 and PO3 will be addressed using multilevel modelling.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Deliverable	**Timelines**
Contract signed	November, 2014
Agreed Protocol for GSK protocol-review forum	November, 2014†
Completion of Statistical analyses plan. Development of programs for analyses.	November 2014 – 1st October 2015
"Data look" SLS data and CPRD data to UoM (Interim 1).	Jan 2016
Analyses of primary objectives (PO1 only) on subset of data using "data look" data (start working on PO2/3)	Feb-March 2016
Final SLS data (one year FU for all subjects) to UoM	1st-8th April 2016
Analyses of PO2 (COPD exacerbation data) using final SLS data: First report with P01/P02 to GSK	By 29th April 2016
Share output from Primary Objectives with SLS Scientific Committee; and CHESS Steering Committee	9-10th May 2016†
First manuscript developed and ready for submission with SLS paper to Thorax based PO1 and PO2 data	June-July 2016†
Programming for secondary objectives and PO3	May-October 2016
100% CPRD-HES data available to UoM (required for PO3 analyses).	Estimated at October 2016
Analysis for PO3 and draft tables circulated	November 2016
Draft complete study report with PO1/PO2/PO3 to GSK	December 2016†
Regulatory reporting of PO3	Q1 2017
Final follow-up manuscript	Q1-2 2017†

† Milestone payment

6. RATIONALE AND BACKGROUND

6.1. Background

Chronic Obstructive Pulmonary Disease (COPD) is a chronic obstructive disease of the airways associated with a significant social and healthcare burden [1, 2, 3]. Most patients with COPD are managed in primary care, as reflected in recent UK guidelines, which are specifically targeted at primary care physicians [4]. The major goals of treatment are to relieve symptoms, improve activity/exercise tolerance, prevent and treat exacerbations, reduce mortality risk and improve health status. However, despite such guidelines, COPD remains under-diagnosed and under-treated; variations in treatments, standards of care and adherence to guidelines have been reported across different geographical regions [2, 5, 6, 7, 8].

Large computerised patient databases provide a useful source of real life observational data, and the General Practice Research Database (GPRD) has been successfully used to generate descriptive epidemiology data in COPD [9, 10, 11, 12] from a large group of UK primary care practices. Historically, the limitations of the GPRD were a time gap between data capture and availability for the researcher and limited links to other healthcare databases, although these are currently being addressed with the development of the Clinical Practice Research Datalink (CPRD) and in recent Phase 4 pragmatic clinical trials [12]. The use of electronic medical record (EMR) data in health research is a key objective in the Department of Health's national research strategy [13].

The Salford Lung Study (SLS) is an ongoing Ph IIIB pragmatic trial comparing a new once daily ICS/LABA fixed dose combination (RELVAR: fluticasone furorate + vilanterol) in which patients are identified by EMR, enrolled in their GP office, randomized to RELVAR or Standard of Care (SOC) maintenance therapy and followed for safety and effectiveness via linked primary care/secondary routine care with primary endpoint of COPD moderate/severe exacerbations over 12 months [14]. MHRA and NICE provided joint advice for the SLS protocol and were supportive of the design to generate "real world" evidence which will demonstrate the value and safety of the medicine against the most relevant standard of care.

6.2. Rationale

Although the SLS will give evidence on the relative effectiveness of RELVAR compared with SOC, the SOC may be prone to the Hawthorne effect, which may distort the effect size.

The Hawthorne effect (also referred to as the observer effect) is a type of reactivity in which individuals improve or modify an aspect of their behaviour in response to their awareness of being observed. The original "Hawthorne effect" studies at the Hawthorne Works in Chicago, USA between 1924 and 1933 suggested that the novelty of being research subjects and the increased attention from such could lead to temporary increases in workers' productivity [15].

In the situation of this study, a potential Hawthorne effect may be a result of potentially different behaviours and decision making of GPs and nurses in their practices of caring for patients with COPD during the SLS study period.

Salford may not be representative of general UK population; hence the prognostic profiles, and potential effect of RELVAR in terms of the outcomes may differ in the general target COPD population compared with Salford.

Both of these issues mean that extrapolation of the results of the SLS to the wider UK population would be subject to major caveats. This proposal aims to explore, and assess how severe the caveats need to be.

7. RESEARCH QUESTION AND OBJECTIVE(S)

Co-primary objectives:

PO1: To characterize the patients enrolled in the Standard of Care (SOC) arm of the SLS compared with the UK population of COPD patients (using the Clinical Practice Research Datalink (CPRD)), including distribution of SES/deprivation level, to evaluate the extent to which the SLS participants are representative of the UK patient population targeted for RELVAR. The comparator set will be specified on two bases: firstly, overall, and secondly, the subset fulfilling the protocol inclusion/exclusion criteria.

PO2: To compare the rate of COPD exacerbation over the 12 months in Standard of Care arm of the SLS compared with the Standard of Care (SOC) recorded in the CPRD, in order to detect a potential Hawthorne effect or other differences.

PO3: To compare the rate of serious pneumonia (defined by hospitalisation) over the 12 months in Standard of Care arm of the SLS compared with the Standard of Care recorded in the CPRD.

Secondary objectives:

SO1: To make comparisons between the SLS Standard of Care and the CPRD cohort on the following health care utilisation (HCU) endpoints: GP visits, hospital admissions mortality and adherence.

SO2: To evaluate other definitions of COPD exacerbation in SOC from CPRD.

SO3: Self-controlled comparison of COPD and other HCU endpoints in Salford before and after SLS commenced.

There are no a-priori hypotheses relating to these objectives.

8. RESEARCH METHODS

8.1. Study Design

This will be an observational COPD cohort study that will utilize the CPRD and the Salford EHR system – to compare selected cohorts with SLS.

For SLS, reference/index date is study entry. This will be matched in CPRD by the following algorithm:

1. Draw up a long-list of potentially eligible individuals in CPRD (patients who would be eligible at some point during the SLS recruitment phase).
2. For each individual:
    a. Randomly sample an entry date from full list of SLS entry dates.
    b. If patient is eligible at that entry date, then they will be included, accounting for relevant immortal time biases up to that entry date, otherwise, they will be excluded. Further details on methods of reducing immortal time bias (i.e left truncated at the entry date and survival modelling) will be outlined in the SAP.

First, the COPD populations in both Salford and in the wider CPRD (excluding Greater Manchester area) will be compared. Second, we will focus on comparisons between all patients enrolled in the SLS SOC arm versus a CPRD cohort that would have been eligible for SLS as per the inclusion and exclusion criteria.

8.2. Setting

In terms of geography, Salford, UK, and surrounding areas within Greater Manchester, for the SLS group; UK-wide, excluding Greater Manchester, for the CPRD group. In terms of health settings, general practice – restricted to practices in Salford and to practices that contribute to CPRD (~10%).

8.2.1. Inclusion Criteria

Two cohorts will be produced. First, a CPRD cohort, using linked primary care, medication, Hospital Episode Statistics, and socio-economic data, according to the following inclusion criteria:

1. Diagnosis of COPD before index date (time period will be defined in the SAP)
2. Aged ≥40 at index date.
3. Alive, and registered with a GP, at index date.
4. Not registered with a GP in the Greater Manchester area.

Second, a Salford cohort will be constructed using the Salford Integrated Record (SIR), according to the inclusion criteria:

1. Diagnosis of COPD before index date.
2. Aged ≥40 at index date.

3. Alive, and registered with a GP, at index date.

Restricted cohorts will then be constructed in both the Salford and CPRD populations, based on the inclusion/exclusion criteria and study period for the SLS:

1. Patients with documented GP diagnosis of COPD, and currently receiving maintenance therapy
2. Male or female subjects aged ≥40 years of age at index date
3. Patients who have a history of treatment with systemic/oral corticosteroids, antibiotics (in association with GP contact) and/or hospitalisation for at least one COPD exacerbation in the 3 years prior to index date.
4. Current COPD Therapy

   All patients currently receiving either:

   - inhaled corticosteroid (ICS) alone or in combination with a long acting bronchodilator (this could be a fixed dose combination or an ICS/LABA provided in two separate inhalers, or ICS and LAMA),
   - or long-acting bronchodilator therapy alone (e.g. tiotropium or salmeterol, or the use of two bronchodilators i.e. LABA/LAMA),
   - or "triple therapy" i.e. ICS/LABA plus a Long Acting Muscarinic Antagonist (LAMA)

Finally, the third data source, the SLS, will be used as-is.

8.2.2. Exclusion Criteria

Subjects meeting any of the following criteria must not be included in the restricted cohorts:

1. Patients with any life threatening condition or uncontrolled/clinically significant disease (code list to be specified in the Study Analysis Plan)
2. Patients with unstable COPD: Patients with an exacerbation (defined by treatment with oral corticosteroids and/or antibiotic or hospital discharge listing COPD) with an onset within 2 weeks of index date. Delay index date until at least 2 weeks after the onset of an exacerbation and until the exacerbation has resolved.
3. Chronic user of oral corticosteroids: Subjects who are considered to be a chronic user of oral corticosteroids for respiratory or other indications (Algorithm to be specified in the Study Analysis Plan).
4. In the Salford population only, those patients who are entered in the SLS and randomised to the RELVAR arm.

8.3. Variables

8.3.1. Outcome definitions

Primary outcomes/endpoint:

- **Rate of COPD exacerbation:** The definition of a COPD exacerbation to be informed by the ongoing study being conducted by Jenny Quint et al. (collaborative project between London School of Hygiene and Tropical Medicine and GSK; GSK study number WEUSKOP5893).

  Moderate/severe COPD exacerbations will be identified using an algorithm combining GP visits, prescriptions for oral corticosteroids and/or antibiotics, or hospital admission, as defined using information from study WEUSKOP5893. Rate of exacerbation during the 12 month follow-up will be calculated and compared with the SLS rate in the standard of care arm; if technically possible, exacerbation rates for the 12 months prior to index date (matched enrolment date) would also be compared.

- **Pneumonia:** To be defined as per the codelist in Table 1 (see 13.1).

Secondary outcomes/endpoints:

- **Healthcare utilisation:** All GP visits/encounters and all hospital admissions during the 12 month study period.
- **Adherence to index prescription:** Defined as percent days covered (PDC) and medication possession ratio (MPR) will also be calculated for the matched cohort, as well as discontinuation, switching medicine or adding on other medicines, to be compared with the SLS SOC arm.
- **Deaths:** All cause, pneumonia death, COPD-attributed death during the 12 month follow-up. For the CPRD, deaths will be determined using Office of National Statistics (ONS) linked mortality data.
- **Other definitions of COPD exacerbation:** Other definitions will be described as per the outputs of study WEUSKOP5893.

8.3.2. Exposure definitions

This is a binary comparison of COPD patients enrolled in the SLS and COPD patients in the CPRD. Hence the primary exposure of interest is whether a patient is enrolled in SLS

(yes/no). A third grouping, COPD patients in Salford (who are not in SLS) will also be examined.

8.3.3. Confounders and effect modifiers

- Sex
- Age
- Socio-economic status (SES)
- Current/SOC COPD Medication group:

  - LAMA or LABA only
  - LAMA+LABA
  - LABA+ICS (combination product or two inhalers)
  - LABA+ICS+LAMA
- Comorbidities
  - Cardio-and cerebrovascular diseases (heart failure, myocardial infarction, stroke)
  - Depression
  - Anxiety
  - Asthma
  - History of pneumonia
  - Gastro-oesophageal reflux and peptic ulcer disease
  - Diabetes
  - Charlson score (COPD will be removed from score)
- Markers of COPD severity
  - Previous COPD exacerbation
  - FEV1 % predicted
  - FEV1/FVC ratio
  - GOLD stage
  - MRC Dyspnoea score
- Comedications: major medication classes for each comorbidity of interest
- Smoking
- BMI
- Vaccinations
- Disability status

Further information on the definitions for the variables above will be provided in the SAP.

8.4. Data sources

The three main data sources are SLS, CPRD and SIR.

The Salford Lung Study (SLS) is described in [14]. In brief, it is a pragmatic trial, carried out in Salford, UK, to evaluate the relative effectiveness and safety of RELVAR compared with SOC. There are two separate studies within the SLS; one for COPD and the other for asthma. For this protocol, the SLS data refers to SLS for COPD only.

The CPRD is a primary care database consisting of a subset of GP practices across the UK. This will be linked to Hospital Episode Statistics (HES), socio-economic status (SES) and Office of National Statistics (ONS) datasets. For brevity, the linked dataset will henceforth be referred to as CPRD.

The SIR is a comprehensive primary and secondary care database detailing healthcare contacts, diagnostic tests and prescriptions of all patients registered with a GP in Salford, UK.

There are a range of subsets and derivations of the data sources that will be considered for this study, listed here for clarity.

CPRD: CPRD – all practices, all COPD patients

CPRD-GM: CPRD – practices/patients in Greater Manchester only, all COPD patients

CPRD-xGM: CPRD – excluding practices/patients in Greater Manchester, all COPD patients

CPRD-GM-IC: CPRD – practices/patients in Greater Manchester only, COPD patients meeting SLS inclusion criteria only

CPRD-xGM-IC: CPRD – excluding practices/patients in Greater Manchester, COPD patients meeting SLS inclusion criteria only

SLS-E: SLS – all eligible. Not all of these are enrolled (some decline)

SLS: SLS – all enrolled

SLS-SOC: SLS – SOC arm only

SIR: SIR – all COPD patients

SIR-IC: SIR - COPD patients meeting SLS inclusion criteria only

In the CPRD, data linkage will be subject to a lag due to the delayed availability of HES and ONS data. Fully linked data will only be available up to a certain date when analyses are undertaken, and as such, primary analyses will be restricted to include SLS enrolled patients up to that date. Subsequent analyses will be conducted once linkage is available for the entire recruitment period. As capture of events of serious pneumonia (as defined in the context of this study) is dependent on records of hospitalisation, PO3 will be analysed when fully linked CPRD-HES data are available.

8.5. Study size

The target sample size for the number of COPD patients enrolled in the SLS is 2,800. Study size in CPRD will be based on the prevalence of COPD diagnosis codes; this will be a minimum 2,800 to match but is expected to be considerably larger (x5-10).

8.6. Data management

8.6.1. Data handling conventions

Data handling within University of Manchester will be governed by the System Level Security Policy for the study (see Annex 1a). Quality control programming will be conducted by within GSK, following data handling SOPs.

8.6.2. Resourcing needs

Staff resources required for the project are outlined in section 2 (responsible parties).

8.6.3. Timings of Assessment during follow-up

As per SLS protocol for SLS patients. CPRD and SIR patients are observational only.

8.7. Data analysis

8.7.1. Essential analysis

All analyses will be conducted using SAS.

For PO1, distributions of the confounders and effect modifiers (as listed in Section 8.3.3) will be tabulated – summarised as proportions in each category for binary and categorical variables, and means/medians and standard deviations for continuous variables. Graphical visualisations will also be produced to aid interpretation (for example, boxplots to characterise age distributions in each population, stacked bar charts to visualise SES by population). This will be done for a series of the derived populations to separate out true differences in demographics in Salford and differences that arise as a consequence of data quality issues etc. The following comparisons will be of interest:

- CPRD-GM v CPRD-xGM: to give an indication of true demographic difference from the same data source.
- CPRD-GM-IC v CPRD-xGM-IC: as above, but restricted to patients meeting the inclusion criteria.
- SIR v CPRD-GM: to give an indication of differences arising as a consequence of selection bias of CPRD practices, and through data quality issues etc.
- SIR-IC v CPRD-GM-IC: as above, but restricted to patients meeting the inclusion criteria.
- SIR-IC v SLS-E: to give an indication of recruitment bias and physician researcher bias (at the approach stage).
- SIR-IC v SLS: to give an indication of recruitment bias (at recruitment stage).
- **SLS v CPRD-xGM-IC**: to indicate the difference between trial recruited, and those meeting the inclusion criteria outside of Salford. This is the key comparison for addressing PO1.

We will then move on to explicit modelling of regional variation of the characteristics within CPRD for the emboldened comparison (**SLS v CPRD-xGM-IC)** to ascertain whether the characteristics observed within Salford are unusual by comparison with CPRD-xGM-IC. We will take local authority regional level (anonymised) as the comparable unit to the study region for SLS. SLS will be considered unusual on a given demographic if an appropriately chosen summary statistic for that demographic (mean for continuous variables) falls outside the 2.5-97.5 percentile range.

For PO2, we will commence with exploratory analyses, similar to described above, to explore the distributions of the primary and secondary endpoints.

Hawthorne effect will be evaluated in two different ways.

Firstly, for descriptive purposes, we will measure the prevalence of the endpoints in a series of subgroups. For example, we will compare the COPD exacerbations in CPRD-xGM-IC with SLS, stratified by SES, gender, etc.

Secondly, we will take a multilevel modelling approach. For this we will combine the SLS and CPRD into one dataset (retaining an indicator of SLS membership). The hierarchies of the model will be patient -> GP practice -> local authority region (with SLS members being treated as a single distinct region) -> strategic health authority region. Strategic health authorities (population threshold of 1 million) are non-anonymised (named) regions. Local authority regions are below the population threshold so an anonymised LA marker will be available.

We will include all confounders and effect modifiers as covariates, with outcomes corresponding to the primary and secondary study outcomes (a separate model for each). Important fixed effects at the local authority level (for example, existence of community teams) will be incorporated into the model if these can be ascertained.

A final model will be selected via backward selection using AIC (Akaike Information Criterion). We will then examine the random effect of the SLS region in the context of the random effects for the other regions. Similar to the above, if the random effect of the SLS region falls outside the 2.5-97.5 percentile range, we will conclude that SOC SLS behaves unusually compared with the rest of the UK, and hence evidence of a Hawthorne effect.

PO3 will be carried out using the same approach as for PO2.

SO1 and SO2 (which pertain to comparing other endpoints, and sensitivity analyses of endpoint definition) will be carried out in the same way as PO2.

SO3 makes explicit the possible change in outcomes at commencement of SLS. We will compare outcome rates within Salford before and after the commencement of SLS, in a self-controlling case design. We will do the same thing within CPRD to control for UK-wide secular trends. This acts as sensitivity analysis to support PO2 (using controls distinct in time rather than in geography).

Additional work will take place exploring the development of methods for a framework of measuring representativeness; this is not within scope of this protocol and is described in Annex 2.1.

8.7.2. Exploratory analysis

8.7.3. General considerations for data analyses

The main identified risks are:

- How linkable the SLS and CPRD datasets are – e.g. can variables be extracted from both with the same (or very similar) definitions for all outcomes and potential confounders.
- Linkage of CPRD to HES and SES is not possible over the calendar dates that SLS has run as there is a lag time until these are released. Hence it may be necessary to restrict some analyses to data from earlier time periods.

8.8. Quality control

CPRD-GOLD has been used previously for descriptive drug utilization studies for prescription medications in respiratory diseases [16, 17, 18].

The standard operating procedures of University of Manchester will guide the conduct of the study, and will include internal quality audits; following rules for secure storage and backup of confidential data and study documentation; quality control procedures for programming, and requirements for senior scientific review.

The QC of analysis will be performed by GSK, in accordance with GSK Standard Operating Procedures (SOPs) and Guidance Documents, specifically the SOP_52213 (4.0) : Conducting Quality Control Review of Worldwide Epidemiology Study Results . The common data model will allow the use of one set of programming following creation of a standardized structure. Wherever feasible, all statistical programming will be independently reviewed by a second analyst, with oversight by a senior statistician. Key study documents, such as the ISAC Protocol, statistical analysis plan, and study reports will undergo quality-control checks and review by the Scientific Steering Committee. Archiving of the project materials will be performed in accordance with GSK SOPs for documentation and archiving of observational studies.

8.9. Limitations of the research methods

Hawthorne effect can only be evaluated for the SOC comparison. This does not give definite evidence about whether the prognostic or predictive effect of RELVAR would differ in the general population. This information could only truly be obtained following use of RELVAR in the general population.

There is no direct metric by which ‘representativeness’ of the Salford cohort can be measured. Early explorations of this will be made in a companion project – see Annex 2.1.

While both the SLS and CPRD use GP data, some data (hospital validated COPD diagnoses, pneumonia data, pharmacy data) for participants in SLS are collected using a different mechanism to CPRD. In this study protocol, only serious pneumonia defined by hospitalization will be assessed, which is a subset of total pneumonia cases recorded in the SLS. Hence any differences (either in representativeness or treatment response) observed between the SLS and non-SLS cohorts could be attributed to differences in data quality and the data collection mechanism. This will be mitigated by an additional comparison of SLS data with CPRD data from within Greater Manchester.

8.9.1. Study closure/uninterpretability of results

Not applicable.

9. PROTECTION OF HUMAN SUBJECTS

9.1. Ethical approval and subject consent

Individual subject consent is not required as this work is using research data. Internal ethical approval will be sought from the University of Manchester.

Linkage of the CPRD to other datasets such as HES is undertaken by a trusted third party (the Health and Social Care Information Centre). The identifiers (date of birth, gender, NHS number, postcode of residence) required for linkage are sent directly from the originating general practice to the trusted third party. CPRD holds only a local patient identifier which is meaningful only at the patients' registered general practice. This identifier is pseudonymised a second time before being made available to researchers and analysts with access to the database.

CPRD's processes have been reviewed by the Confidentiality Advisory Group (CAG) and approved by the Health Research Authority (HRA) and Secretary of State to process patient identifiable information without consent under Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002. This effectively removes the obligation to obtain patient consent for the use of confidential patient information for conducting purely observational research using CPRD databases, and associated linked datasets. This approval is conditional on approval of a study protocol by the CPRD Independent Scientific Advisory Committee (ISAC). In addition to ISAC approval, the protocol will be reviewed by GSK Worldwide Epidemiology Protocol Review Forum.

9.2. Subject confidentiality

The SLS data will be anonymised at source by the SLS team, before this is passed to University of Manchester.

The CPRD only contains fully de-identified patient data. No patient identifiable information will be available to the study team, or to GSK. All data held and processed by CPRD and any other study partners will be done so in compliance with the relevant legal obligations including the Data Protection Act 1998.

All data will be held on a secure computer network, with access restricted to authorised users.

10. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

This is a retrospective study. From the CPRD, free text data will not be available to allow causality determination of any potential adverse events. Adverse events arising from the SLS trial will have previously been reported appropriately during the trial period.

11. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

11.1. Target Audience

This work is targeted 1) internally at GSK, 2) regulators and 3) the wider scientific community; in order to understand the SLS in wider context. Results will be disseminated externally primarily by manuscripts.

11.2. Study reporting and publications

Reporting and publications according to the following table:

Deliverable	**Timelines**
Agreed Protocol for GSK protocol-review forum	November, 2014
Completion of Statistical analyses plan. Development of programs for analyses.	November 2014 – 1st October 2015
Analyses of PO2 (COPD exacerbation data) using final SLS data: First report with PO1/PO2 to GSK	By 29th April 2016
Share output from Primary Objectives with SLS Scientific Committee; and CHESS Steering Committee	9-10th May 2016
First manuscript developed and ready for submission with SLS paper to Thorax based PO1 and PO2 data	June-July 2016
Analysis for P03 and draft tables circulated	November 2016
Draft complete study report with PO1/PO2/PO3 to GSK	December 2016
Regulatory reporting of PO3	Q1 2017

Final follow-up manuscript	Q1-2 2017

In addition, we will present the results of the study at international respiratory conferences as appropriate. The study protocol and results will be posted to GSK Clinical Study Register as per GSK SOPs.

11.3. Data Sharing

For SLS SOC data: These data fall under the GSK SHaring Anonymised REsearch Data (SHARE) initiative; researchers are may request data via a committee approval process.

For the CPRD data: any CPRD license holder may request the data used in this study via the usual CPRD ISAC process.

The study protocol will available on the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENcEPP) website (http://www.encepp.eu/). Therefore, the study can be replicated by any interested third party.

12. REFERENCES

1. Vestbo J, Hurd SS, Agusti AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M, Stockley RA, Sin DD, Rodriguez-Roisin R. Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease, GOLD Executive Summary. Am J Respir Crit Care Med 2013, 187: 347-365.
2. Mathers CD, Loncar D. Projections of Global Mortality and Burden of Disease from 2002 to 2030. PLoS Med 2006, 3(11): e442.
3. Lopez AD, Shibuya K, Rao C, Mathers CD, Hansell AL. Held LS, Schmid V, Buist S. Chronic obstructive pulmonary disease: current burden and future projections. Eur Respir J 2006, 27: 397–412.
4. Bellamy D, Bouchard J, Henrichsen S, Johansson G, Langhammer A, Reid J, van Weel C, Buist S. International Primary Care Respiratory Group (IPCRG) Guidelines: Management of Chronic Obstructive Pulmonary Disease (COPD). Prim Care Respir 2006, 15:48-57.
5. Sobradillo-Peña V, Miravitlles M, Gabriel R, Jiménez-Ruiz CA, Villasante C, Masa JF, Viejo JL, Fernández-Fau L. Geographic variations in prevalence and underdiagnosis of COPD: results of the IBERPOC multicentre epidemiological study. Chest 2000, 118: 981-9.
6. Soriano JB, Miravitlles M, Borderías L, Duran-Tauleria E, García Río F, Martínez J, Montemayor T, Muñoz L, Piñeiro L, Sánchez G, Serra J, Soler-Cataluña JJ, Torres

A, Viejo JL, Sobradillo-Peña V, Ancochea J. Geographical Variations in the Prevalence of COPD in Spain: Relationship to Smoking, Death Rates and other Determining Factors. Arch Bronconeumol. 2010, 46: 522-530.

7. Aisanov Z, Bai C, Bauerle O, Colodenco FD, Feldman C, Hashimoto S, Jardim J, Lai CKW, Laniado-Laborin R, Nadeau G, Sayiner A, Shim JJ, Tsai YH, Walters RD, Waterer G. Primary care physician perceptions on the diagnosis and management of chronic obstructive pulmonary disease in diverse regions of the world. Int J Chron Obstruct Pulmon Dis 2012, 7: 271–282.
8. Corrado A, Rossi A. How far is real life from COPD therapy guidelines? An Italian observational study. Respir Med. 2012, 106: 989-97.
9. Hansell A, Hollowell J, Nichols T, McNiece R, Strachan D. Use of the General Practice Research Database (GPRD) for respiratory epidemiology: a comparison with the 4th Morbidity Survey in General Practice (MSGP4). Thorax 1999, 54: 413–419.
10. Soriano JB, Maier WC, Egger P, Visick G, Thakrar B, Sykes J, Pride NB. Recent trends of physician-diagnosed COPD in women and men in the UK. Thorax 2000, 55: 789–794.
11. Hansell AL, Lam KA, Richardson S, Visick G, Soriano JB. Medical event profiling of COPD patients. Pharmacoepidemiol Drug Saf 2004, 13: 547–555.
12. van Staa T-P, Goldacre B, Gulliford M, Cassell J, Pirmohamed M, Taweel A, Delaney B, Smeeth L. Pragmatic randomised trials using routine electronic health records: putting them to the test. BMJ 2012, 344: e55 doi: 10.1136.
13. Department of Health (Research and Development Directorate) policy paper: Best research for best health: a new national health research strategy. 25 January 2006. Last accessed 06 June 2013 www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4127127.
14. New JP, Bakerly ND, Leather D, Woodcock A. Obtaining real-world evidence: the Salford Lung Study. Thorax. 2014 Apr 26 e pub ahead of print.
15. French J.R.P. Experiments in field settings. In: Festinger L., Katz D., editors. Research methods in the behavioral sciences. Holt, Rinehart & Winston; New York, NY: 1953
16. DiSantostefano RL, Sampson T, Le HV, Hinds D, Davis KJ, Bakerly ND. Risk of Pneumonia with Inhaled Corticosteroid versus Long-Acting Bronchodilator Regimens in Chronic Obstructive Pulmonary Disease: A New-User Cohort Study. PLoS One. 2014; 9(5):e97149.
17. Ashworth M, Latinovic R, Charlton J, Cox K, Rowlands G, Gulliford M. Why has antibiotic prescribing for respiratory illness declined in primary care? A longitudinal study using the General Practice Research Database. *J Public Health (Oxf)*. 2004; 26(3):268-274.
18. van Staa TP, Cooper C, Leufkens HG, Lammers JW, Suissa S. The use of inhaled corticosteroids in the United Kingdom and the Netherlands. Respir Med. 2003; 97(5):578-585.

ANNEX 1. LIST OF STAND-ALONE DOCUMENTS

Annex 1a: System Level Security Policy (SLSP) for Study

System Details

1. **The System shall be known as**

   CHESS: CPRD-COPD Hawthorne Effect Study in Salford: A UK cohort study to characterise patients enrolled in the Salford Lung Study and to evaluate a potential Hawthorne effect

2. **The System's responsible owner shall be** Matthew Sperrin

3. **The System's Caldicott Guardian or Data Controller shall be** Matthew Sperrin

System Security

4. **Security of the system shall be governed by the corporate security policy of University of Manchester**

   http://documents.manchester.ac.uk/display.aspx?DocID=6525 (policy)

   http://documents.manchester.ac.uk/display.aspx?DocID=8039 (responsibilities)

5. **The System's responsible security manager shall be:**

   Tony Arnold, University IT Security Coordinator

6. **The security manager duties shall include**:

   Devise, implement, enforce and review the University's IT security and data handling policies.
   Being the first point of contact for any security related queries or concerns.
   Being consulted on and providing the final sign-off for any requests for change to any aspects of IT security for the system.

7. **The System shall incorporate the following security countermeasures**:

   - **Physical Security – Data Processing**: The researchers are based within Vaughan House which is swipe card access only from reception into the building. Staff and Postgraduate students must have their University swipe cards enabled for access to the building. The offices are also locked when vacant.
   - **Physical Security – Data Hosting**: The data will be stored within Personal Drives (P: Drives) hosted on the University's network storage infrastructure which is the recommended location for storing sensitive or critical University

data. The storage infrastructure is hosted across two data centres (approx. 2KM apart) for resilience and disaster recovery purposes. Physical access to the data centres is strictly limited to data centre staff and a limited number of authorised IT Services staff. The data centres are protected by physical and electronic access security systems, swipe card access in and out of the data centres and CCTV coverage. The data centres are locked down out of hours and access is discouraged, but can be arranged by prior agreement with the data centre manager.

- **Access Control and Privilege Management:** The data will be hosted on authorised system user's P: drives which are strictly controlled data shares within the University's network storage infrastructure to which only the owner of the P: Drive has access permissions. The data share will only be accessed via a mapped network drive from PCs identified for research data processing.
- **Network Security Measures:** Network access control lists prevent PCs outside of the campus LAN from accessing the network storage infrastructure.
- **Other – Data Processing:** Four PCs have been identified for processing the research data. Once the PC has loaded the operating system a local, password protected computer account is required to login to the PC. This account is unique to the primary user of the computer and only the account owner knows the password. The PC has the Windows firewall enabled and configured to prevent remote access. The PCs have been configured to automatically update their antivirus signatures daily and have been configured to download and install any Microsoft operating system and application security patches automatically from the Microsoft update service.

System Management

8. **The System shall be developed / provided by:**

   University of Manchester, Faculty of Medical & Human Sciences, Information Services
   University of Manchester, IT Services Division

9. **The System shall be implemented & maintained by:**

   University of Manchester, Faculty of Medical & Human Sciences, Information Services will configure and maintain the security aspects of the PC, user accounts and access controls for the data share on the network storage infrastructure.

   University of Manchester, IT Services Division (ITSD) will be responsible for providing secure, reliable data hosting on the network storage infrastructure.

   Servers procured by ITSD include maintenance on either 3 or 5 year agreements depending on the Service requirements. Supplier engineers replace any defectives items and may request access to the Data Centre. In addition Data centre staff are trained and able to carry out component replacements on behalf of the suppliers.

Storage arrays are procured with support and maintenance included as part of a 3 year package. The SAN Arrays in the central ITS Data Centres are supplied by EMC and have allocated engineers who are familiar with our site configuration and conversant in maintaining the equipment and advising on future changes.

To ensure the security of University hosted infrastructure systems, all system changes must be authorised via the change management process. Any proposed system changes are recorded as requests for change (RFC's) and authorised by the change advisory board (CAB).

10. **The System shall be shared or used by the following organisations:**

GSK will have direct access to the system for quality checking purposes.

System Design

11. **The System shall comprise:**

The research PCs connect to the University's network via access switches which are located in data cabinets within secure, dedicated comms rooms. The switches are logically segregated into separate VLAN's for network efficiency and security. The access switches then connect to the University's core router and onto perimeter routers via multiple paths for resilient access to the data centres where the network storage infrastructure is hosted. The perimeter routers connect onto the JANET network and the wider internet. The perimeter routers are configured with access control lists which provide security for incoming network traffic. A network diagram can be found at the end of this SLSP.

The operating system on the PCs identified for data processing require local username and password authentication for access. The P: drives on the network require username and password authentication also.

Operational Processes

12. **The patient identifiable / sensitive data will be collected:**

Datasets will be pseudonymised at source by providers. No patient identifiable or sensitive information will be processed.

13. **The data will be stored:**

The data will be stored electronically in R and SAS file formats. The data will be made available to authorised members of research staff via P: drives hosted on the University's network storage infrastructure housed in the University's data centres.

14. **The data will be processed**:

Four University approved desktop PCs will be used to process the data. The PCs will not cache copies of the data and all data will be stored on the network storage infrastructure.

The University's Information Handling Policy sets out how digital information should be handled. This includes confidentiality, integrity and availability and the use of encryption tools for the protection of sensitive information and communications.

http://www.itservices.manchester.ac.uk/medialibrary/pdf/secureguidance/GP-InformationHandling.pdf

15. **The System's authorised users shall be** :

University of Manchester:
- Matthew Sperrin
- Tjeerd Van Staa
- Jane Candlish
- 1 x Research Associate, to be appointed.

GSK
- TBA by GSK.

The system's authorised users are all University members of staff and individuals designated by GSK. The data will not be accessible by any other third party organisations

16. **When the system or its data has completed its purpose / has become redundant or is no longer needed, the following methods will be adopted to dispose of equipment, back-up media or other stored data**:

Sensitive material on removable media are deleted as soon as possible. Printed materials and CD/DVDs containing sensitive information are shredded when no longer required. When the analysis is completed the researcher will delete files. All items of equipment containing storage media shall be checked to ensure that any sensitive data and licensed software has been removed or securely overwritten.

Desktop PCs are disposed of when replaced via a recognised disposal company, Computer Disposals LTD (CDL). CDL erase the hard drives to Government Restricted Standard SEAP (UK), which is three overwrites plus an additional verification pass. A certificate is produced for every successfully data erased hard drive to include the make, model and serial number of the hard drive. Any hard drive that fails the data erase process is degaussed on site at CDL using the latest CESG approved degausses and forwarded for recycling.

University IT Services has a policy of securely wiping network storage infrastructure arrays onsite prior to disposal. Disks are securely erased by software aligned to the DoD5220-22M standard and are then disposed of via CDL who also wipe the disks as per their procedure outlined above.

System Audit

17. **The System shall benefit from the following internal / external audit arrangements:**

In 2006, the University's IT Services undertook (in conjunction with KPMG) a comprehensive IT Risk Management Benchmarking exercise to appraise the University's approach to the management of IT-related risks. The review was intended to provide a benchmark upon which to build the maturity of IT Services activities over a number of years. Whilst the exercise found a number of positive areas within IT, it also identified a number of areas for development and agreed management action plans to address the issues raised.

A follow-up review was conducted during 2009 by UNIAC, the University's internal auditing body. This review: (i) revisited each original recommendation, ascertaining progress to date (supported by testing where appropriate) and its ongoing relevance; (ii) commented on the adequacy of the actions to date; (iii) proposed revised action plans for previous actions remaining outstanding; (iv) made additional suggestions over and above the agreed action.

The original KPMG report contained over thirty high and medium level recommendations and, realistically, a number of them would take a considerable period to fully implement. The follow-up review in 2009 concluded that: (i) recommendations had been fully implemented with no further action required in seven areas; (ii) good progress had been made towards implementing a further fifteen recommendations; (iii) some progress had been made towards implementing a further six recommendations; (iv) limited progress had been made towards implementing a final four recommendations.

Overall, the report concluded that progress has been encouraging and indicated that IT Services management had provided adequate focus to improving the management of IT risks.

Future auditing arrangements include regular audits agreed with internal and external auditors. The Director of IT Services meets the UNIAC Director annually to agree the internal programme; the external programme is agreed via the University's Audit Committee.

18. **The System shall be risk assessed every 12 months**

The University's Compliance and Risk Officer (CRO) is responsible for ensuring that the University is meeting its many statutory and regulatory compliance obligations. The CRO is responsible for supporting the University's risk management process, all aspects of risk management and has developed a risk management framework.

All major University functional areas (including IT Services) are required to conduct annual risk assessments and to review risk registers on a quarterly basis. Risk registers are submitted to the University's CRO for reporting to the

University's Risk Management Committee. Risk management is a specific responsibility of heads of operational areas.

System Protection

19. **The System shall benefit from the following resilience / contingency / disaster recovery arrangements:**

    The University's storage infrastructure is hosted and replicated across two data centres (approx. 2KM apart) for resilience and disaster recovery purposes.

    The University's IT Services Division (ITSD) utilises Legato Networker Backup domains. Supporting infrastructure comprises disk libraries and both physical and virtual tape libraries. Cross data centre backup is performed, so services hosted within data centre 1 (Kilburn) are backed up to data centre 2 (Reynold) and vice versa.

    Backup/recovery plans are documented as part of the service install process during the commissioning of a specific service. Each Service is responsible for its business continuity and disaster recovery plans, to which ITSD feed in its technical recovery plans

    ITSD operates a change management process. All proposed changes to infrastructure hosted, maintained and administered by ITSD are recorded via the RFC process with changes being authorised by a Change Advisory Board (CAB).

20. **In the event of serious disruption or total system failure, business continuity shall be provided by the following means:**

    The University's geographically dispersed, replicated, twin data centre approach with cross site backup has been designed to be as fault tolerant as possible and to provide business continuity in the event of a data centre failure. Should a situation arise where both data centres became unavailable then the University's disaster recovery plans relating to the failed system would be implemented.

21. **In the event of a security or confidentiality breach occurring the following procedure shall be followed:**

    Information on the procedure for reporting a security or confidentiality breach is available from the following link on the University's Secure-IT website: http://www.itservices.manchester.ac.uk/secure-it/reporting/

SSP Ownership

22. **This SLSP shall be the responsibility of:**

    Matthew Sperrin

22.1 - Shall be reviewed on an annual basis for its completeness and for relevant update.

23. **The SLSP shall be available / distributed to:**

    Authorised GSK and University of Manchester staff involved in research activities or members of IT staff assisting with the completion of SLSP forms.

- Through which secure means:

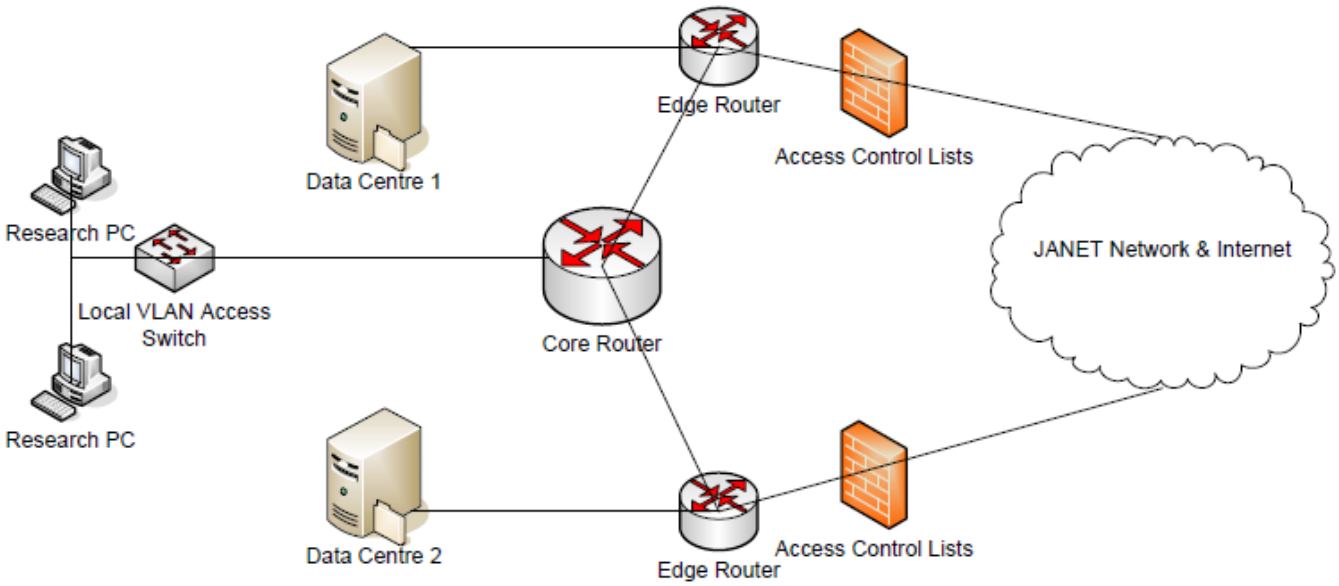
    The SLSP document will be distributed to authorised GSK and University of Manchester staff via the internal email system.

Data Protection Registration

24. **Please confirm that your organisation has Data Protection Registration to cover the purposes of analysis and for the classes of data requested.**

    http://www.ico.gov.uk/ESDWebPages/search.asp - . Registration No: Z6787610

Network diagram as referenced in the System Design section of the SLSP



ANNEX 2. ADDITIONAL INFORMATION

A2.1 Development of methods for the measurements of representativeness

This part of the project will deal with the development of methods for measuring representativeness of trials and evaluating potential effects of non-representativeness. It will be explorative and focused on methods. The concept of representativeness of trials is widely known but it is less known how to actually measure representativeness. This project will evaluate RELVAR use as an exemplar case study. These results will be compared to a historic case study of selective Cox-2 inhibitors (comparing the registration trials); the Cox-2 analyses will be conducted and funded as part of the GetReal IMI project. The RELVAR project will include the following activities:

(i) Review of literature for methods that can measure level of representativeness and evaluate the effects of non-representativeness. These methods may include multilevel models (including levels of clinician, practice, patient, disease and exposure characteristics.

(ii) Risk prediction models will be developed in CPRD for the outcomes of interest (to be defined). This analysis will determine the risk factors for the outcomes of interest. In addition, experts will be asked to provide likely effect modifiers of RELVAR. The analyses will focus on risk factors and effect modifiers.

(iii) Three populations will be identified in CPRD:

a. the first population will be based on the inclusion and exclusion criteria of the SLS trial. In case that information is missing in the EHR, methods will be evaluated to possibly impute these criteria.

b. the second population will be based on expert views of the likely possible use of RELVAR, and will be compared to Cox-2 inhibitors (from GetReal) in actual clinical practice. As an example, the trials for selective Cox-2 inhibitors were conducted in narrowly defined populations while later used in very broad population (replacing traditional NSAIDs). Any analysis of representativeness would have showed a considerable difference between the populations potentially eligible for a trial and potential users in actual clinical practice.

c. the third population will consist of all COPD patients aged 40 years or older alive at the index date. The propensity score for recruitment into SLS (as based on the Salford data) will be applied to this population, estimating the probability that a patient could have been recruited into SLS.

(iv) The analyses will include comparisons of the distribution of risk factors and effect modifiers between these three populations and the trial populations. Also, a comparison of incidence rates for the outcomes of interest will be conducted across these populations. Methods will be developed to integrate these results.

This project will be conducted in collaboration with GetReal partners, including NICE.

13. TABLES

13.1. Table 1: Pneumonia definition/codes used in SLS

Within the SLS, these codes are used assess pneumonia from hospital discharge records.

ICD-10 code	ICD-10 description	Comment
B67.1	Other B67.1 Echinococcus granulosus infection of lung	Other
J17.3	Other J17.3 Pneumonia in parasitic diseases	Other
J16	Pneumonia due to other infectious organisms NEC	unspecified
J16.8	Pneumonia due to other specified infectious organisms	unspecified
J17	Pneumonia in diseases classified elsewhere	unspecified
J17.8	Pneumonia in other diseases classified elsewhere	unspecified
J18	Pneumoniaorganism unspecified	unspecified
J18.0	Bronchopneumonia, unspecified	unspecified
J18.1	Lobar pneumonia, unspecified	unspecified
J18.8	Other pneumonia, organism unspecified	unspecified
J18.9	Pneumonia, unspecified	unspecified
A06.5	Amoebic lung abscess	Lung abscess
J85	Abscess of lung and mediastinum	Lung abscess
J85.0	Gangrene and necrosis of lung	Lung abscess
J85.1	Abscess of lung with pneumonia	Lung abscess
J85.2	Abscess of lung without pneumonia	Lung abscess
B20.6	HIV disease resulting in Pneumocystis carinii pneumonia	Fungal
B37.1	Pulmonary candidiasis	Fungal
B38.0	Acute pulmonary coccidioidomycosis	Fungal
B38.1	Chronic pulmonary coccidioidomycosis	Fungal
B38.2	Pulmonary coccidioidomycosis, unspecified	Fungal
B39.0	Acute pulmonary histoplasmosis capsulation	Fungal
B39.2	Pulmonary histoplasmosis capsulati, unspecified	Fungal
B40.0	Acute pulmonary blastomycosis	Fungal
B40.2	Pulmonary blastomycosis, unspecified	Fungal
B41.0	Pulmonary paracoccidioidomycosis	Fungal
B42.0	Pulmonary sporotrichosis	Fungal
B44.0	Invasive pulmonary aspergillosis	Fungal

B44.1	Other pulmonary aspergillosis	Fungal
B45.0	Pulmonary cryptococcosis	Fungal
B46.0	Pulmonary mucormycosis	Fungal
B58.3	Pulmonary toxoplasmosis	Fungal
B59.X	Pneumocystosis	Fungal
J17.2	Pneumonia in mycoses	Fungal
A15	Respiratory TB bacteriologically and histologically confirmed	Mycobacterial
A15.0	TB lung confirm sputum microscopy with or without culture	Mycobacterial
A15.1	Tuberculosis of lung, confirmed by culture only	Mycobacterial
A15.2	Tuberculosis of lung, confirmed histologically	Mycobacterial
A15.3	Tuberculosis of lung, confirmed by unspecified means	Mycobacterial
A15.4	TB intrathoracic lymph nodes confirm bact histologically	Mycobacterial
A15.5	Tuberculosis of larynx, trachea & bronchus conf bact/hist'y	Mycobacterial
A15.6	Tuberculous pleurisy, conf bacteriologically/his'y	Mycobacterial
A15.7	Primary respiratory TB confirm bact and histologically	Mycobacterial
A15.8	Other respiratory TB confirm bact and histologically	Mycobacterial
A15.9	Respiratory TB unspec confirm bact and histologically	Mycobacterial
A16	Respiratory TB not confirmed bacteriologically or histologically	Mycobacterial
A16.0	Tuberculosis of lung, bacteriologically & histolog'y neg	Mycobacterial
A16.1	Tuberculosis lung bact and histological examin not done	Mycobacterial
A16.2	TB lung without mention of bact or histological confirm	Mycobacterial
A16.5	TB pleurisy without mention of bact or histological confirm	Mycobacterial
A16.7	Prim respiratory TB without mention of bact or hist confirm	Mycobacterial
A16.8	Oth respiratory TB without mention of bact or hist confirm	Mycobacterial
A16.9	Resp TB unspec without mention of bact or hist confirm	Mycobacterial
A19	Miliary tuberculosis	Mycobacterial
A19.0	Acute miliary tuberculosis of a single specified site	Mycobacterial

A19.1	Acute miliary tuberculosis of multiple sites	Mycobacterial
A19.2	Acute miliary tuberculosis, unspecified	Mycobacterial
A19.8	Other miliary tuberculosis	Mycobacterial
A19.9	Miliary tuberculosis, unspecified	Mycobacterial
A31.0	Pulmonary mycobacterial infection	Mycobacterial
B01.2	Varicella pneumonia	Viral
B05.2	Measles complicated by pneumonia	Viral
J10.0	Influenza with pneumonia, influenza virus identified	Viral
J11.0	Influenza with pneumonia, virus not identified	Viral
J12	Viral pneumonia, not elsewhere classified	Viral
J12.0	Adenoviral pneumonia	Viral
J12.1	Respiratory syncytial virus pneumonia	Viral
J12.2	Parainfluenza virus pneumonia	Viral
J12.8	Other viral pneumonia	Viral
J12.9	Viral pneumonia, unspecified	Viral
J17.1	Pneumonia in viral diseases classified elsewhere	Viral
A20.2	Pneumonic plague	Bacterial
A21.2	Pulmonary tularaemia	Bacterial
A22.1	Pulmonary anthrax	Bacterial
A42.0	Pulmonary actinomycosis	Bacterial
A43.0	Pulmonary nocardiosis	Bacterial
A48.1	Legionnaires' disease	Bacterial
J13	Pneumonia due to Streptococcus pneumoniae	Bacterial
J13.0	Pneumonia due to Streptococcus pneumoniae	Bacterial
J13X	Pneumonia due to Streptococcus pneumoniae	Bacterial
J14	Pneumonia due to Haemophilus influenzae	Bacterial
J14.0	Pneumonia due to Haemophilus influenzae	Bacterial
J14X	Pneumonia due to Haemophilus influenzae	Bacterial
J15	Bacterial pneumonianot elsewhere classified	Bacterial
J15.0	Pneumonia due to Klebsiella pneumoniae	Bacterial
J15.1	Pneumonia due to Pseudomonas	Bacterial
J15.2	Pneumonia due to staphylococcus	Bacterial
J15.3	Pneumonia due to streptococcus, group	Bacterial

	B	
J15.4	Pneumonia due to other streptococci	Bacterial
J15.5	Pneumonia due to Escherichia coli	Bacterial
J15.6	Pneumonia due to other aerobic Gram-negative bacteria	Bacterial
J15.7	Pneumonia due to Mycoplasma pneumoniae	Bacterial
J15.8	Other bacterial pneumonia	Bacterial
J15.9	Bacterial pneumonia, unspecified	Bacterial
J16.0	Chlamydial pneumonia	Bacterial
J17.0	Pneumonia in bacterial diseases classified elsewhere	Bacterial
B25.0	Cytomegaloviral pneumonitis	No
B38	Coccidioidomycosis	No
B38.9	Coccidioidomycosis, unspecified	No
B39	Histoplasmosis	No
B39.4	Histoplasmosis capsulati, unspecified	No
B39.5	Histoplasmosis duboisii	No
B39.9	Histoplasmosis, unspecified	No
B40	Blastomycosis	No
B409	Blastomycosis, unspecified	No
B44	Aspergillosis	No
B44.9	Aspergillosis, unspecified	No
J18.2	Hypostatic pneumonia, unspecified	No
J86.0	Pyothorax with fistula	No
J86.9 Pyothorax without fistula	Pyothorax without fistula	No
'X' denotes that all subcodes under the 3-digit main number are included		
"No" in category denotes not assigned to a major category		