TITLE PAGE

Division: Worldwide Development
Retention Category: GRS019
Information Type: Worldwide Epidemiology Study Protocol

Title:	Can social listening data be used to provide meaningful insights into abuse or inappropriate use of bupropion? (A feasibility analysis)

Compound Number:	GR67205
Development Phase	IV
Effective Date:	24 June 2015
Subject:	Drug abuse in relation to drug safety
Author(s):	[redacted]

PASS information

Title	**Can social listening data be used to provide meaningful insights into potential abuse or inappropriate use of bupropion? (A feasibility analysis)**
Protocol version identifier	02
Date of last version of protocol	12/12/14

EU PAS register number	Registered 20 Jan 2015
Active substance	GR67205 [List of pharmacotherapeutic group(s) (ACT codes) and active substance(s) subject to the study]
Medicinal product	Wellbutrin, Wellbutrin XL, Wellbutrin SR, Zyban
Product reference	GR67205
Procedure number	[If applicable, Agency or national procedure number(s), e.g. EMA/X/X/XXX]
Marketing authorisation holder(s)	Glaxo Wellcome UK Stockley Park West Uxbridge Middlesex UB11 1BT
Joint PASS	No
Research question and objectives	**Can social listening data be used to provide meaningful insights into potential abuse or inappropriate use of bupropion?** Objectives: 1. To determine if social media can identify cases of potential abuse or inappropriate use of bupropion which can effectively complement existing sources of data currently used for pharmacovigilance activities 2. To explore the utility of three internet forums to identify cases of interest 3. To describe and characterize the posts of interest (POI) identified during this feasibility analysis (POI is a term coined to denote a user post that may be relevant to drug abuse or inappropriate use of the product in question)

Country(-ies) of study	Worldwide (internet data from English-speaking sites, from much of which no geographic data will be available)
Author	[redacted] MD, RPh, FACOG Contracting Safety Physician, PVCOI GCSP RTP, NC 27709 [redacted]

MARKETING AUTHORISATION HOLDER(S)

Marketing authorisation holder(s)	Glaxo Wellcome UK Stockley Park West Uxbridge Middlesex UB11 1BT UK
MAH contact person	[redacted] contracting safety physician on behalf of GSK

1. TABLE OF CONTENTS

PAGE

2. LIST OF ABBREVIATIONS

AE	Adverse Event
API	Application Programming Interfaces
CRaWL	Contextualizing ReAl World use of drugs through social Listening, a project sponsored by the Pharmacovigilance Center of Innovation
CSD	Central Safety Department
GSK	GlaxoSmithKline
HCPs	Health Care Professionals
PII	Personally Identifiable Information
POI	"Post of interest", a term coined to denote a user post that may be relevant to drug abuse or inappropriate use of the product in question
RSS Feeds	Rich site summary feeds

3. RESPONSIBLE PARTIES

- [redacted] principle investigator, SERM Manager, CMO GCSP Mature Products US
- [redacted], SERM Director, CMO GCSP SERM US
- [redacted], Contracting Safety Physician, Pharmacovigilance Center of Innovation
- [redacted], SERM Director, CMO GCSP Mature Products US and Pharmacovigilance Center of Innovation
- [redacted], Chief Data Scientist, Epidemico™
- [redacted], Epidemiology Director, Pharma R and D
- [redacted], SERM Manager, CMO GCSP SERM UK
- [redacted] SERM Medical Director, CMO GCSP SERM UK
- [redacted], Medical Advisor, NS, Classic and Established Medicines
- [redacted], US Medical Affairs

WWEpi Project number:

SPONSOR SIGNATORY:

[redacted]MD, RPh, FACOG
Primary Author of Protocol/ Project officer

15.6.2015
Date

[redacted]Head, [redacted]M Development
Therapy Area Leader

16.6.2015
Date

SPONSOR INFORMATION PAGE

WWEpi Project Identifier: eTrack Study Number 202115

Sponsor Legal Registered Address:

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

Sponsor Contact Address

GlaxoSmithKline Research & Development Limited
Five Moore Drive
P.O. 13398
Research Triangle Park, NC 27709-3398, USA
Telephone: [redacted]

In some countries, the clinical trial sponsor may be the local GlaxoSmithKline affiliate company (or designee). Where applicable, the details of the Sponsor and contact person will be provided to the relevant regulatory authority as part of the clinical trial submission.

Sponsor Medical Monitor Contact Information: n/a (noninterventional)

4. ABSTRACT

Title:

Rationale and background:

Purpose: to determine the feasibility of using social media for collecting meaningful insights into potential abuse or inappropriate use of bupropion.

Rationale: From proof of concept evaluations for Project CRaWL (Contextualizing Real World drug use through social Listening), we know that information about abuse or inappropriate use potential of marketed GSK drugs is available through social listening (the process of identifying and assessing what is being said about a company, individual, product or brand on the Internet). The quality and quantity of those data are not fully explored at this time or in any formal evaluation setting.

These data are typically quite scant in the standard tools used for pharmacovigilance (spontaneous adverse event reporting, observational databases, and literature reports) as abusers are unlikely to report adverse events or means of abuse to regulatory authorities or often even their personal physician. In a recent review of data from the DAWN database, it was reported that, "There are several limitations to the data used in the study, which preclude the ability to make strong recommendations as to the abuse potential of bupropion." [Bibeau,2012] (The data, however, did not provide evidence that abuse of bupropion was growing.)

Knowing that there are some data available from online forums and even mainstream social media sites, we believe that further exploration of these data may be useful. In this feasibility study for bupropion as an example drug, we hope to describe the best use of the data collection tool that we are using through a partnership with Epidemico™, an informatics company with interest and experience in this realm.

Background

Bupropion hydrochloride was first approved in the US in December of 1985 and is currently approved in 80 countries, for depression, smoking cessation, and for seasonal affective disorder. It is classified as a substance of low abuse potential [Miller, 1983; Griffity, 1990; Rush, 1998; Zernig, 2004].

Abuse of bupropion has been described in published case reports, and was first recognized in the setting of correctional institutions where illicit drugs are less available and where bupropion may be widely ordered as a smoking cessation therapy for prison inmates. Most of these reports involve routes other than oral use (the only approved route of administration) including nasal insufflations and intravenous injection [Kim, 2010; Barribeau, 2013; Hilliard, 2013; Reeves, 2013; Yoon, 2013]. During a recent search of the DAWN database there was a paucity of data on route of administration or confirmation of psychoactive effects [Bibeau, 2012]. This is one potential area where social listening data may help significantly augment the existing sources of information that we have on bupropion's abuse or inappropriate use potential.

Research question and Objectives:

- To determine if social media can identify cases of potential abuse or inappropriate use of bupropion which can effectively complement existing sources of data currently used for pharmacovigilance activities
- To explore the utility of three internet forums to identify cases of interest
- To describe and characterize the posts of interest (POI) identified during this feasibility analysis

Study Design:

This is a retrospective descriptive observational study, analyzing all available data collected on bupropion and comparator drugs (venlafaxine and amitriptyline, as noted in section 7) from three internet forums known to be rich with drug abuse data. Summary statistics on numbers of posts, threads, and authors for drugs of known high abuse potential will also be collected and provided for contextualization of the bupropion and comparator data.

Population:

Data, from publically available social media or internet forum posts from individuals who choose to post on a number of sites will be collected by Epidemico[tm] through the DataSift[TM] platform or directly from the in-scope website administrators. The population will thus be self-selecting and voluntary, and may include users from any country or background as long as they post in the English language and agree to the site's policies.

Variables: The data recorded will include:

- Number of posts of interest (POI), a term coined to denote a post that describes or is related to the abuse or inappropriate use of a drug in question, identified over the study period
- Demographic data where available: age, gender, geographic location, education level/occupation, race/ethnicity
- Number of total posts needed to identify a POI
- Indicator scores for POI vs posts of non-interest
- Site-specific and population-specific results of above endpoints
- IMS sales data for North America and Europe

Data sources:

All data are to be provided to GSK by Epidemico after application of Epidemico's automated classifying software (see section 9 for further description). Only publically available internet data with NO PII (personal identifying information) will be provided to GSK. Websites to be searched will include bluelight.org, erowid.org, and opiophile.org.

Study size:

All available posts from the above sites will be included. The number of unique patients or posts collected is unknown prior to actual data collection.

Data analysis:

The data obtained will be manually reviewed by GSK's Safety Listening Lab team members with both clinical expertise and experience with the manual curation process for pertinence to drug abuse or inappropriate use of bupropion and comparators. Specific endpoints of interest will include route of administration, dosage and length of use, categorization of euphoric effect, whether prison/criminal justice system is involved, procurement of the drug, magnitude of abuse problem within a community, and combination with other agents (polypharmacy). Outputs will then be graphically displayed and comment will be made on the feasibility of this data collection and reporting method as a tool to enhance current pharmacovigilance efforts. Additionally, inclusion of some exemplary verbatim posts in the final report will help the audience conceptualize the tool and dataset.

Milestones:

Key data collection timelines (timepoints) are as follows:

Day One (TBD after SRT and PRF approval)—Retrospective data received from Epidemico

Day 30— Randomly selected 200 posts reviewed, team meets to discuss initial findings or need for further protocol amendment

Day 120—iterative results reviewed by team and next steps discussed

5. AMENDMENTS AND UPDATES

Amendment or update no	Date	Section of study protocol	Amendment or update	Reason
1	31 Mar 2015	4, 6, 7, 9	• Prospective data collection changed to retrospective • In-scope websites changed and thus wording of objective # 2 changed • Project CRaWL team name changed to Safety Listening Lab • Comparator drug handling changed • No longer focusing only on possible Proto-AEs, but curating all posts • Variable for collection added: Mention of magnitude of the abuse problem within the community	• Learned from direct contact with sites that retrospective data are available and would lead to quicker results • Initial data from planned websites offered no new insights • Project CRaWL concluded, led to creation of Safety Listening Lab March 2015 • Increasing experience with resources needed for manual curation led team to feel more strongly about focusing on unanswered questions and not re-visiting abuse potential of drugs of known high potential • Very little data on performance of the automated classifier in this realm led to team concerns about missing valuable information by relying on the classifier tool to filter the data for us • From Section 9.6.1 below After initial data exploration, new data points may need to be added in order to record unforeseen points.
<2>	<Date>	<Text>	<Text>	<Text>
<n>	<Date>	<Text>	<Text>	<Text>

6. MILESTONES

Milestone	Planned date
Start of data collection	30 May 2015
End of data collection	30 July 2015
<Study progress report 1>	30 September 2015
<Study progress report 2>	
<Study progress report n>	
<Interim report 1>	<Date>
<Interim report 2>	<Date>

<Interim report n>	<Date>
Registration in the EU PAS register	20 Jan 2015
Registration in eTrack (#202115)	6 Nov 2014
SRT Approval	14 Nov 2014, amendment 1 = 6 May 2015
PRF Approval	12 Dec 2014, amendment 1=
Final report of study results	30 December 2016

7. RATIONALE AND BACKGROUND

Abuse potential, as defined in FDA's draft guidance document for Assessment of Abuse Potential of Drugs [FDA, 2010], refers to a drug that is used in nonmedical situations, repeatedly or even sporadically, for the positive psychoactive effects it produces. These drugs are characterized by their central nervous system (CNS) activity. Examples of the psychoactive effects they produced include sedation, euphoria, perceptual and other cognitive distortions, hallucinations, and mood changes. Drugs with abuse potential often (but not always) produce psychic or physical dependence and may lead to the disorder of *addiction.* The concept of *abuse potential* encompasses all the properties of a drug, including, for example, chemical, pharmacological, and pharmacokinetic characteristics, as well as fads in usage and diversion history.

Specifically, the FDA states in the draft guidance that, "sponsors should search publicly available databases, including the Drug Abuse Warning Network (DAWN), the National Survey on Drug Use and Health (NSDUH), the Treatment Episode Data Set (TEDS), Monitoring the Future (MTF), and *other databases* [emphasis is protocol author's], to characterize and monitor risks associated with the misuses and abuse of a drug and to estimate the extent of use and abuse of a particular drug."

The draft guidance also includes suggestions for collecting information on the type of data that might be collected for a product (eg abuse or inappropriate use events as "numerator" and amount of drug produced in the same time period as "denominator") that could be compared with information on pharmacologically similar drugs. The FDA also suggests that such data could be used to evaluate trends over time.

Background on bupropion abuse [from PRJ2215]:

In early preclinical studies, bupropion showed amphetamine-like effects in animals. Drug discrimination studies in rodents and primates indicate that the subjective experience (stimulus cue) elicited by bupropion is generalized to stimulants such as d-amphetamine, cocaine, and methylphenidate [de la Garza, 1987; Bergman, 1989;Kamien, 1989; Lamb, 1990] .

Despite the evidence for stimulant effects in animals, suggesting a relevant abuse or inappropriate use potential, several clinical studies in humans indicated that oral intake of bupropion had lower abuse liability than amphetamine, methylphenidate, and even caffeine. Accordingly, it was concluded that bupropion did not exhibit amphetamine-like

characteristics in humans, and the drug has been classified as a substance of low abuse potential [Miller, 1983; Griffity, 1990; Rush, 1998; Zernig, 2004].

Abuse potential had been part of the Benefit Risk Management Plan for bupropion up until 2003 and at that point, had no longer been regarded as a potential risk that required additional/further evaluation outside standard pharmacovigilance monitoring. The current European Risk Management Plan also states that standard pharmacovigilance monitoring applies to abuse potential. Routine pharmacovigilance monitoring during 2013 identified an increase in the number of spontaneous reports of bupropion abuse in the GSK worldwide safety database (OCEANS).

The Bupropion Safety Review Team (SRT) agreed that although the numbers of abuse reports were small relative to the total number of reports for bupropion in OCEANS, there was sufficient information to warrant investigation of the potential effect on public health. PRJ2215 was performed to evaluate the route of administration of bupropion resulting in abuse or misuse of the drug in the DAWN database. There were several limitations to the data used in the study, which precluded the ability to make strong recommendations as to the abuse potential of bupropion. However, the data from this study "did not provide evidence that abuse and misuse of bupropion is growing." [Bibeau, 2012]

The SRT is interested in additional sources of data that may help inform the abuse or inappropriate use potential and real-time abuse of bupropion that may be derived from the proposed study. We plan to use social media listening to better understand bupropion abuse potential.

Background on Social Listening and The Safety Listening Lab Currently, post-marketing safety surveillance relies on data from spontaneous adverse event reports, published literature and observational databases (medical records, insurance claims). These data sources have limitations that include: significant under-reporting (some estimate that less than 10% of adverse events are reported), lack of geographically diverse data (most data come from the United States and Europe), and time lag (most data sources lag 9-12 months).

Social listening, a term used to describe the process of monitoring social media data, is widely used in many industries (and governments) and this led GSK's Central Safety Department (CSD) to ask if social listening could be leveraged for pharmacovigilance. Project CRAWL (**C**ontextualizing **ReA**l **W**orld drug use through social **L**istening), a pilot to evaluate the benefits and risks of using social listening for post-marketing safety surveillance, was recently launched to evaluate this technology for routine pharmacovigilance. The evaluation comprised a set of research studies on the data that the tool can provide as well as two pilot projects using the data for GSK drug monitoring. Results from Project CRaWL led to the development of the Safety Listening Lab within the Pharmacovigilance Center of Innovation at GSK. The application of this tool to evaluate abuse concerns for medications for which we are actively seeking new data sources is one of the first projects of the Safety Listening Lab, spelled out in this protocol.

8. RESEARCH QUESTION AND OBJECTIVE(S)

QUESTION: **Can social listening data be used to provide meaningful insights into potential abuse or inappropriate use of bupropion? (A feasibility analysis)**

Purpose

- The purpose of this analysis is to determine the feasibility of using social media for collecting meaningful insights into potential abuse or inappropriate use of bupropion.

Objectives and Endpoints

Objectives	Measured Outcomes
• To determine if social media can identify cases of potential abuse or inappropriate use of bupropion which can effectively complement existing sources of data currently used for pharmacovigilance activities	• Number of posts of interest (POI)* identified over a period of time • Total number of posts that must be reviewed in order to identify each POI* • Describe indicator scores for POI* vs non-interest posts
• To explore the utility of three internet forums to identify cases of interest	• Site-specific results of above endpoints • Population-specific results of above endpoints
• To describe and characterize the posts of interest (POI) identified during this feasibility analysis	• Descriptive data

*POI (post of interest) is a post that describes or is related to the abuse or inappropriate use of bupropion. This may be better defined at the end of the feasibility portion of the study. Of note, posts will be reviewed manually by reviewers blinded to site/source in order to guard against introduction of bias here.

9. RESEARCH METHODS

9.1. Study Design

This is a non-traditional feasibility study design using a novel data source which we are collaborating with the informatics company Epidemico[tm] to apply to pharmacovigilance. The design is essentially a retrospective descriptive observational study. Data will be collected retrospectively from internet websites and forums where drug abuse or inappropriate use may be discussed and voluntarily posted on a public site. Data will be collected on chatter concerning bupropion, as well as other drugs that are known to have lower or similar abuse potentials to bupropion (venlafaxine and amitriptyline). Summary statistics on numbers of posts, threads, and authors concerning drugs of known high abuse potential (oxycodone, buprenorphine, methylphenidate, and alprazolam) will also be reported to help contextualize the data.

9.1. Setting

Setting: Public internet forums where drug abuse is discussed, including bluelight.org, erowid.org, and opiophile.org. Posts in the English language will be included in our search, and we will collect all posts mentioning bupropion products as well as comparator products (venlafaxine, amitriptyline).

Data will be collected using DataSift™, a commercial social media/Big Data collection and delivery service (see below) and/or in direct cooperation with the website owners. Epidemico™ will then provide their commercially available deidentified data to GSK. The medical product data are acquired from publically available online forums that are accessible through proprietary automated content scraping technology, Application Programming Interfaces (APIs) officially published by the sources/sites, and RSS feeds.

After data are acquired, they undergo classification by Epidemico – a filtering process in which an automated Bayesian classifier removes irrelevant items (including duplicates and spam) and further categorizes the language presented in the data. Using the same conceptual process as spam filters for email, the classifier has been trained with a machine learning algorithm to recognize language that may describe an adverse event. The classifier then uses a proprietary vernacular-to-regulatory dictionary to translate symptoms described in colloquial and slang terms into MedDRA terminology (e.g., "skin looks like a lobster" and "I look like a beet" would be classified to erythema).

An indicator score is thus assigned by Epidemico's software developed for this application, ranging from 0 to 1 and indicating the machine-derived likelihood that the post is related to any adverse event. . Manual review of the posts done by GSK's Safety Listening Lab team members with both clinical expertise and experience with the manual curation process will then inform the cut-off level for this score in potential future applications of the product.

About DataSift (from DataSift website) [DataSift, 2014]

DataSift Inc. is the platform that powers the social economy, enabling companies to aggregate, filter and extract insights from the billions of public social conversations on Twitter, leading social networks and millions of other sources. DataSift provides access to both real-time and historical social data to uncover insights and trends that relate to brands, businesses, financial markets, news and public opinion. Key investors include Insight Venture Partners, Scale Venture Partners, Upfront Ventures and IA Ventures. DataSift has offices in San Francisco, New York City and Reading, U.K.

9.2. Variables

All posts mentioning bupropion products as well as comparators (and common misspellings and slang terms) will be acquired:

- All posts on the in-scope internet sites will be reviewed, as the sites are specifically targeted to chatter concerning abuse or inappropriate use
- All posts will be categorized into

- Abuse-related
- Proto-AEs (above indicator score thresholds)
- Otherwise meaningful mention (zero to indicator score thresholds)
- Unclear/uncodable or spam (negative indicator scores)

Epidemico's automated classifier will also be applied, which is designed to flag posts where product tampering is mentioned (e.g., for purposes of injecting, snorting, etc), or where any misuse, abuse, or diversion is likely. Bupropion classifier posts will then be manually curated by GSK HCPs with the Safety Listening Lab to remove false positives and to request reclassification of mistakes made by the automated classifier. The manually curated bupropion abuse-related posts will then be described in the following settings. Individual POI may be presented in one or more categories described below as appropriate:

- Route of Administration – through manual review. Reviewers will note if the following were mentioned:
  - Nasal insufflations (e.g., snorting)
  - Oral- chew
  - Oral- swallow
  - Smoking
  - Intravenous
  - Injection
  - Subcutaneous
  - Ambiguous and other routes of administration – internet jargon (during initial feasibility project, will determine possibility of further differentiation amongst routes of administration)

- Dosage and length of use

- Categorization of euphoric effect – for all posts identified as abuse-related, the nature of the high will be broadly characterized as being (example terms that would be mapped to the characterizations follow each term):
  - Stimulant-like ("upper")—CNS stimulation, insomnia, energy/energized, increased heart rate, decreased appetite, seizures, increased confidence, excitement, rush, nervousness, anxiety, anger, euphoria
  - CNS depressant-like ("downer")—sedative, anxiolytic, nerve pill, tranquil/tranquilizer, CNS depression, slowed heart rate, slow respiration, sleep/drowsiness, dull senses, diminished pain, slurred speech, coma, hypnotic
  - Other dissociative effects and hallucination ("all arounder")—psychedelic, distorted perceptions, nausea, dizziness, sweating, raised blood pressure, distorted sensory messages, illusion, altered perception, intensified external stimulus perception, delusions, delirium

- Unknown or unspecified

- Prison/criminal justice flag
  - Given the nature of the case reports, any interaction with the criminal justice system (prison, jails), etc. will be manually flagged using a dichotomous indicator.

- Procurement comments—drug prescribed for patient, obtained/purchased illegally from street/market, obtained/stolen from family member or other acquaintance

- Polypharmacy: extraction of the names of other substances ingested simultaneously or in combination with bupropion

- IMS Sales data for North America and Europe

- Demographic information where available: age, gender, geographic location, education level/occupation, and race/ethnicity

- Mention of magnitude of the abuse problem within the community

9.2.1. Outcome definitions

- "Drug abuse, dependence and withdrawal" SMQ (see Annex One)
  - Combined category --
    - Abuse
    - Misuse
    - Dependence
    - Overdose
    - Diversion

9.2.2. Exposure definitions

Exposures for capture will include all posts mentioning bupropion or any of the comparator drugs in the English language.

9.2.3. Confounders and effect modifiers

Confounding and effect modification will not be explored in this non-traditional and non-interventional descriptive study.

9.3. Data sources

See also section 9.2, Setting

Several public internet forums where drug abuse is discussed, including bluelight.org, erowid.org, and opiophile.org. Posts in the English language will be included in our search, and we will collect all posts mentioning bupropion products as well as comparator products (venlafaxine, amitriptyline). Summary statistics will be collected for comparators of high abuse potential also as noted in section 9.1

9.4. Study size

There are no a priori specified hypotheses for this study which would drive sample size calculations. All eligible cases will be included, and we will manually curate all cases obtained for bupropion and comparator drugs venlafaxine and amitriptyline.

9.5. Data management

See also Section 9.2, setting.

Data will be collected by Epidemico using DataSift[tm], a commercial social media/Big Data collection and delivery service (see below), as well as directly from website owners where applicable/necessary. Epidemico[tm] will then ensure the data are deidentified and cleared of all personal identifiable information (PII) before performing automated classification and providing to GSK. Epidemico's medical product data are acquired from online forums that are accessible through proprietary automated content scraping technology, Application Programming Interfaces (APIs) officially published by the sources/sites, and RSS feeds.

After data are acquired, they undergo classification by Epidemico as above – a filtering process in which an automated Bayesian classifier removes irrelevant items (including duplicates and spam) and further categorizes the language presented in the data. Using the same conceptual process as spam filters for email, the classifier has been trained with a machine learning algorithm to recognize language that may describe an adverse event. The classifier then uses a proprietary vernacular-to-regulatory dictionary to translate symptoms described in colloquial and slang terms into MedDRAterminology (e.g., "skin looks like a lobster" and "I look like a beet" would be classified to erythema).

An indicator score is thus assigned by Epidemico's software developed for this, ranging from 0 to 1 and indicating the machine-derived likelihood that the post is related to drug abuse. Manual review of the posts will then inform the cut-off level for this score in future applications of the product and real-time use for monitoring bupropion abuse internet chatter.

The automated classifier/indicator score is designed to flag posts where product tampering is mentioned (e.g., for purposes of injecting, snorting, etc), or where any misuse, abuse, or diversion is likely. The manual curation process will lead to review of all abuse-related posts to extract further information.

Manual curation will be conducted by GSK Health Care Providers with specific expertise and used to remove false positives and to request reclassification of mistakes made by

automated classifier. A custom interface or standard spreadsheet software may be used for this. Findings can then be fed back in to the process to better inform the machine "learning" and improve the tool.

9.5.1. Data handling conventions

See above. No direct patient cases for handling and all data deidentified for PII before becoming available to GSK.

9.5.2. Resourcing needs

We estimate that approximately 1 FTE for 4-6 weeks will be needed for data analysis, interpretation and reporting. Money will also be spent on data acquisition. These expenses will be covered by the budget for the Safety Listening Lab, sponsored by [redacted]

9.5.3. Timings of Assessment during follow-up

First assessment 30 days after full dataset available, as noted in section 4 above. .

9.6. Data analysis

9.6.1. Essential analysis

Objectives
• To determine if social media can identify cases of potential abuse or inappropriate use of bupropion which can effectively complement existing sources of data currently used for pharmacovigilance activities
• To explore the utility of three internet forums to identify cases of interest
• To describe and characterize the posts of interest (POI) identified during this feasibility analysis

Objective One: (To determine if social media can identify cases of potential abuse of bupropion which can effectively complement existing sources of data currently used for pharmacovigilance activities)

Data on all reports of bupropion related to abuse will be captured and examined. Data from each post will be extracted and descriptive statistics reported in a summary table as available. (See example table one, Annex One).

After initial data exploration, new data points may need to be added in order to record unforeseen points (such as whether the drug was noted to be for primary or additive effect in the case of significant polypharmacy, whether there is a new term for a drug mentioned that we had not known to look for previously, whether vernacular-to-regulatory mapping was appropriate or needs to be adjusted, or other variables)

Bupropion results will be presented alongside the results of comparator drugs in order to provide some context around frequency of mention and relative public health burden.

Formal comparisons of abuse potential across drugs are not possible given the limitations of the data (please see Limitations, Section 9.9). In an effort to adjust for availability/circulation of the drugs, the number of abuse mentions per unit sold will be calculated for each product using available IMS sales data.

Summative graphs can then be used for visual data description as seen in Annex One.

Objective Two: (To explore the utility of three internetforums to identify cases of interest) Data will also be described and graphically displayed based on the site or forum from which it was gleaned (which the manual curators will be blinded to prior to data analysis). Since no a priori definitions of site utility exist in this space, only descriptive reports can be provided.

Objective Three: (To describe and characterize the posts of interest (POI) identified during this feasibility analysis)

Qualitative description of the data will then be reported. Inclusion of some exemplary verbatim posts in the final report will help the audience conceptualize the tool and dataset. Word-mapping or other contextualization tools may also be applied to better understand and describe the chatter about bupropion abuse.

9.6.2. Exploratory analysis

- These will be driven by the essential analyses, but if the sample size permits, we will report the data by route of administration, dosage and length of use, categorization of euphoric effect, prison/criminal justice flag, procurement comments, polypharmacy, IMS sales data, demographic information, and mention of magnitude of the abuse problem within the community to help us understand the strengths and weaknesses of this tool.

9.7. Quality control

This is an original design with data collection for a novel purpose via a novel method. There is no prior validation. This is the reason for including the comparator drugs as well as IMS sales data for denominator comparisons, consistent with the FDA's draft guidance document [FDA, 2010].

9.8. Limitations of the research methods

The current study is primarily designed to assess the utility of social media in detecting a signal of abuse potential of a product.

There are currently some major limitations in the use of social media information in terms of quantitative signal evaluation where the abuse potential of one product is

compared to the abuse potential of another product. These limitations relate to missing information for the numerator and unclear denominators for any comparisons of abuse frequency.

Regarding the numerator, mention frequency is likely to be driven by access to the medication (if relatively low availability/ circulation of a drug, the potential for diversion or misuse will be lower). Although we intend to provide some context around the number of bupropion mentions versus other drugs by assessing mentions/ unit sold using IMS sales data as a proxy for availability/circulation, there are limitations to this approach. Sales data may not be available for all countries and there may be access controls applied to a product in some which will not be reflected in the sales data.

Mention frequency may also be affected by the type of experience that the abuse/misuse results in. For example, it may be that more dramatic effects/ experiences ("highs" or psychedelic experiences) are mentioned more frequently than effects such as somnolence. If products differ in the type of experience they elicit, comparisons may be biased due to differential potential for missing information in the numerator.

The extent of information missing from those who abuse/misuse a substance and do not post about it cannot be evaluated. Furthermore, the demographic profile of those who post about their substance abuse/misuse compared to the profile of those who abuse-misuse substances but don't post is relatively unknown. Validation studies which compare the demographic profile of those who abuse/misuse and post versus those who abuse/misuse a product and don't post are likely to be difficult to conduct given the difficulty in identifying those who abuse/misuse substances and the legal implications involved. If a product that is more likely to be abused/misused by individuals who don't post about their experience is compared to a product that is more likely to be abused/misused by people who do post, then these comparisons will be biased.

Other limitations of studying this novel tool include: difficulty mapping vernacular terminology to standard regulatory dictionaries, ever-changing and evolving nature of vernacular speech, and confounding by spam and advertisement-type posts as well as the imperfect nature of the deduplication tools for posts.

9.8.1. Study closure/uninterpretability of results

If data volume is not sufficient for analysis after data collection, we will consider termination of this project.

10. PROTECTION OF HUMAN SUBJECTS

10.1. Ethical approval and subject consent

Informed consents not applicable in this collection of data offered willingly to public internet forums by self-selected patients. Data will be collected only from sites that permit it by their user agreements and ethics approval is not required.

10.2. Subject confidentiality

All data in this study are publically-available and deidentified as part of Epidemico's standard commercial product offerings prior to being provided to GSK. The study team is working with patient privacy experts to ensure that this is protected to the best of our ability and will flag any potential concerns back to Epidemico for continuous quality improvement of the data collection and deidentification system.

11. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

During this study, reportable adverse events will not be noted due to the nature of the deidentified data. The following governance has been put in to place:

- For our social media listening project, we will be purchasing de-identified data from a third party vendor that has been stripped of Personally Identifiable Identification. Therefore, in the absence of an identifiable reporter, we will have no individual case reporting requirements. We will instead report any signals either in an expedited manner or as part of routine aggregate reports in keeping with how we currently treat observational data from other sources.
- In order to ensure alignment and acceptance both internally and externally, we have already consulted and communicated this approach with the FDA, MHRA and EMA, GSB, the OCMO Leadership Team, Global Digital Risk Board, the patient privacy office, PV compliance, regulatory compliance, legal, IT, the joint GCSP/GRA leadership team, and others.

12. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

12.1. Target Audience

We will aim to disseminate these results in the form of a peer-reviewed journal article at the end of the project. GSK stakeholders will have the opportunity to review the information generated by the study prior to submission for publication, including the bupropion clinical team, GCSP and Safety Listening Lab team and sponsors, and any other applicable or interested parties.

12.2. Study reporting and publications

Upon protocol approval, protocol summary will be posted to both the EU PAS Register and the GSK Clinical Study Register (VCTR).

Upon completion of the study, results will be posted on the GSK Clinical Study Register, the EU PAS Register, and will be prepared in manuscript form for journal submission.

13. REFERENCES

Bergman J, Madras B, Johnson SE and Spealman RD. Effects of cocaine and related drugs in nonhuman primates. III. Self-administration by squirrel monkeys. J Pharmacol Exp Ther 1989; 251: 150-5.

Bibeau KB, Henegar C. Assessment of bupropion misuse and abuse 2004-2011. GSK Clinical Study Result Summary http://www.gsk-clinicalstudyregister. com/study/201235#ps Accessed 18 Nov 2014.

DataSift, Press release: Recognized for innovation in social data. http://datasift.com/ press-releases/ dataSift-recognized-for-innovation-in-social-data/ Accessed 25 Nov 2014.

de la Garza R, Johanson CE. Discriminative stimulus properties of intragastrically administered d-amphetamine and pentobarbital in rhesus monkeys. J Pharmacol Exp Ther. 1987;243:955–962

FDA, DRAFT Guidance for Industry, Assessment of Abuse Potential of Drugs, www.fda.gov_downloads_drugs_guidancecomplianceregulatoryinformation_guidances_ucm198650, Accessed 20 Nov 2014.

Griffith JD, Carranza J, Griffith C, et al. Bupropion: clinical assay for amphetamine-like abuse potential. J Clin Psychiatry. 1983;44.

Kamien JB and Woolverton. A pharmacological analysis of the discriminative stimulus properties of d-amphetamine in rhesus monkeys. J Pharmacol Exp Ther 1989; 248: 938-46.

Lamb RJ and Griffiths RR. Self-administration in baboons and the discriminative stimulus effects in rats of bupropion, nomifensine, diclofensine and imipramine. Psychopharmacology 1990; 102: 183-90.

McNaughton EC, Black RA, Zulueta MG, Budman, SH, Butler SF. Measuring online endorsement of prescription opioids abuse: an integrative methodology. Pharmacoepidemiology and drug safety 2012; 21: 1081–1092

Miller L and Griffith J. A comparison of bupropion, dextroamphetamine, and placebo in mixed-substance abusers. Psychopharmacology. 1983;80:199-205

Rush CR, Kollins SH, Pazzaglia PJ. Discriminative-stimulus and participant-rated effects of methylphenidate, bupropion, and triazolam in D-amphetamine-trained humans. Exp Clin Psychopharmacol 1998.

Shutler L, Perron J, Portelli I, Nelson LS, Blachford CR. Prescription opioids in the twittersphere: A contextual analysis of tweets about prescription drugs. Annals of Emergency Medicine 2013;62(4S):S122.

Zernig G, De WH, Telser S, et al. Subjective effects of slow-release bupropion versus caffeine as determined in a quasi-naturalistic setting. Pharmacology 2004;70.

ANNEX 1. LIST OF STAND-ALONE DOCUMENTS

Tables

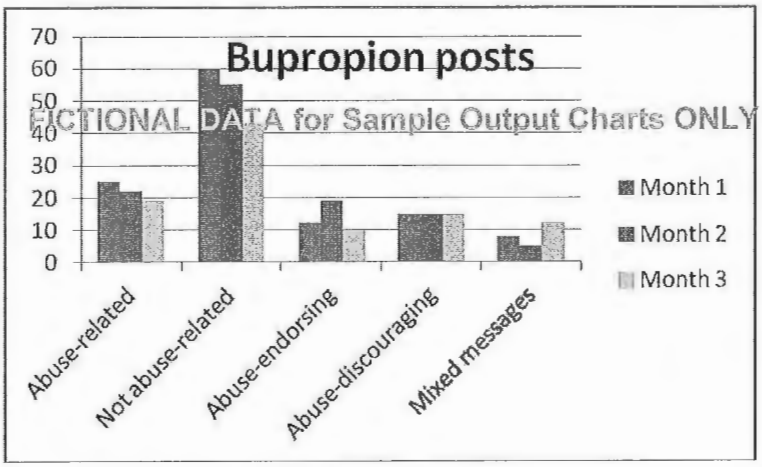
Example table one: Descriptive characteristics of bupropion data

	Bupropion (n)	Bupropion (%)	Venlafaxine controls (n)	Venlafaxine controls (%)	Amitriptyline (n)	Amitriptyline(%)
Number of drug mentions		100%		100%		100%
Number of abuse-related mentions						
Route of Administration: Oral- Chewed						
Oral- swallowed						
Nasal						
Smoking						
Intravenous						
Subcutaneous						
Injection not otherwise classified						
Ambiguous or other						
Categorization of Euphoric Effect: Stimulant-like/"upper"						
CNS-depressant-like/"downer"						
Other dissociative effects or hallucination/"all-						

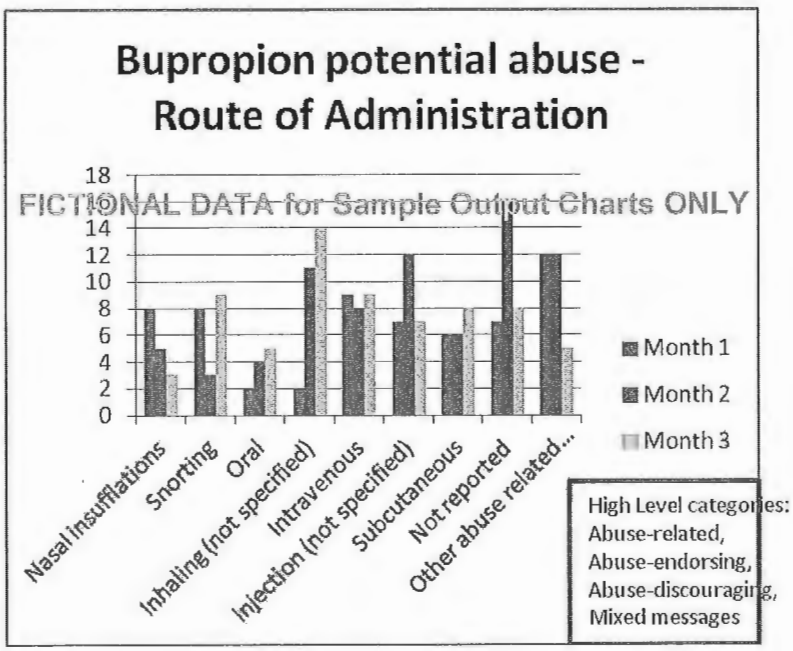
arounder"						
Unknown or unspecified						
Prison or criminal justice involvement?						
Procurement information available?						
Prescribed for patient?						
Obtained illegally (buy/trade)						
Obtained/stolen from family member of acquaintance?						
Polypharmacy/concomitant drugs used?						
Mention of magnitude of abuse problem in community?						

Figures, for example purposes only. Captions will change based on final data availability.

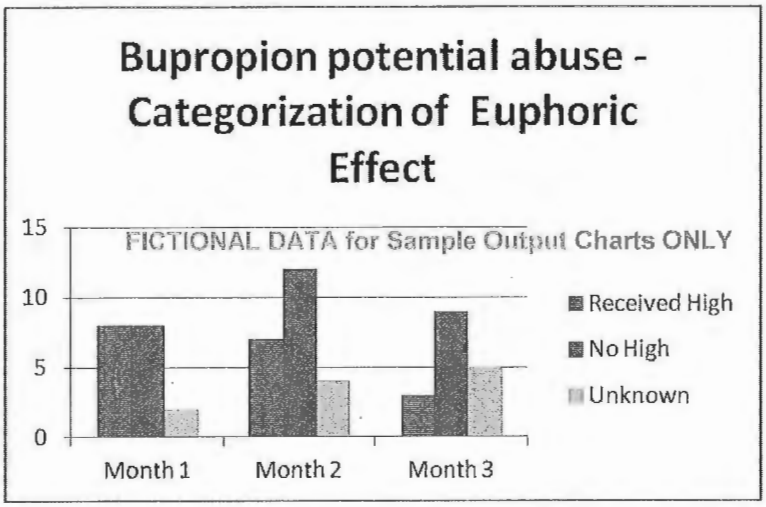
Bupropion – High Level Categories



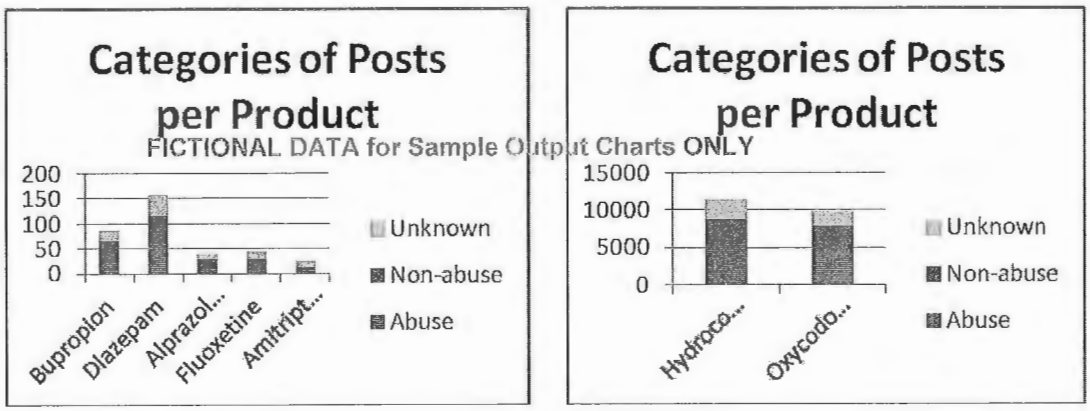
Bupropion potential abuse – Route of Administration



Bupropion potential abuse – Categorization of Euphoric Effect



Categories of Posts per Product



FICTIONAL DATA for Sample Output Charts ONLY

No.	Document Reference No	Date	Title
1.	1	18 Nov 2014	SMQ for drug abuse
2.	<No>	<Date>	<Text>
N	<No>	<Date>	<text>

PT	PT code	SMQ
Dopamine dysregulation syndrome	10067468	Drug abuse and dependence
Drug abuse	10013654	Drug abuse and dependence
Drug abuser	10061111	Drug abuse and dependence
Drug dependence	10013663	Drug abuse and dependence
Drug dependence, antepartum	10013675	Drug abuse and dependence
Drug dependence, postpartum	10013676	Drug abuse and dependence
Intentional drug misuse	10065679	Drug abuse and dependence
Intentional overdose	10022523	Drug abuse and dependence
Maternal use of illicit drugs	10026938	Drug abuse and dependence
Neonatal complications of substance abuse	10061862	Drug abuse and dependence
Polysubstance dependence	10053243	Drug abuse and dependence
Substance abuse	10066169	Drug abuse and dependence
Substance abuser	10067688	Drug abuse and dependence
Accidental overdose	10000381	Drug abuse and dependence
Dependence	10012335	Drug abuse and dependence
Disturbance in social behaviour	10061108	Drug abuse and dependence
Drug detoxification	10052237	Drug abuse and dependence
Drug diversion	10066053	Drug abuse and dependence
Drug level above therapeutic	10061132	Drug abuse and dependence
Drug level increased	10013722	Drug abuse and dependence
Drug screen	10050837	Drug abuse and dependence
Drug screen positive	10049177	Drug abuse and dependence
Drug tolerance	10052804	Drug abuse and dependence
Drug tolerance decreased	10052805	Drug abuse and dependence
Drug tolerance increased	10052806	Drug abuse and dependence
Medication overuse headache	10072720	Drug abuse and dependence
Narcotic bowel syndrome	10072286	Drug abuse and dependence
Needle track marks	10028896	Drug abuse and dependence
Overdose	10033295	Drug abuse and dependence
Prescription form tampering	10067669	Drug abuse and dependence
Substance use	10070964	Drug abuse and dependence
Substance-induced mood disorder	10072387	Drug abuse and dependence
Substance-induced psychotic disorder	10072388	Drug abuse and dependence
Toxicity to various agents	10070863	Drug abuse and dependence
Drug withdrawal convulsions	10013752	Drug withdrawal
Drug withdrawal headache	10013753	Drug withdrawal
Drug withdrawal maintenance therapy	10052970	Drug withdrawal
Drug withdrawal syndrome	10013754	Drug withdrawal
Drug withdrawal syndrome neonatal	10013756	Drug withdrawal
Drug rehabilitation	10064773	Drug withdrawal
Rebound effect	10038001	Drug withdrawal

Steroid withdrawal syndrome	10042028	Drug withdrawal
Withdrawal arrhythmia	10047997	Drug withdrawal
Withdrawal syndrome	10048010	Drug withdrawal
Device defective	10074425	Lack of efficacy/effect
Device failure	10056871	Lack of efficacy/effect
Device ineffective	10059875	Lack of efficacy/effect
Drug effect decreased	10013678	Lack of efficacy/effect
Drug effect delayed	10068303	Lack of efficacy/effect
Drug effect incomplete	10013682	Lack of efficacy/effect
Drug effect variable	10074541	Lack of efficacy/effect
Drug half-life reduced	10049994	Lack of efficacy/effect
Drug ineffective	10013709	Lack of efficacy/effect
Drug ineffective for unapproved indication	10051118	Lack of efficacy/effect
Drug level decreased	10013718	Lack of efficacy/effect
Drug resistance	10059866	Lack of efficacy/effect
Drug specific antibody present	10013745	Lack of efficacy/effect
Drug tolerance	10052804	Lack of efficacy/effect
Drug tolerance increased	10052806	Lack of efficacy/effect
Multiple-drug resistance	10048723	Lack of efficacy/effect
No therapeutic response	10063670	Lack of efficacy/effect
Paradoxical drug reaction	10048958	Lack of efficacy/effect
Tachyphylaxis	10043087	Lack of efficacy/effect
Therapeutic product ineffective	10060769	Lack of efficacy/effect
Therapeutic product ineffective for unapproved indication	10060770	Lack of efficacy/effect
Therapeutic reaction time decreased	10061380	Lack of efficacy/effect
Therapeutic response decreased	10043414	Lack of efficacy/effect
Therapeutic response delayed	10053181	Lack of efficacy/effect
Treatment failure	10066901	Lack of efficacy/effect
Vaccination failure	10046862	Lack of efficacy/effect
Virologic failure	10065648	Lack of efficacy/effect
Drug withdrawal convulsions	10013752	Drug withdrawal
Drug withdrawal headache	10013753	Drug withdrawal
Drug withdrawal maintenance therapy	10052970	Drug withdrawal
Drug withdrawal syndrome	10013754	Drug withdrawal
Drug withdrawal syndrome neonatal	10013756	Drug withdrawal
Drug rehabilitation	10064773	Drug withdrawal
Rebound effect	10038001	Drug withdrawal
Steroid withdrawal syndrome	10042028	Drug withdrawal
Withdrawal arrhythmia	10047997	Drug withdrawal
Withdrawal syndrome	10048010	Drug withdrawal

ANNEX 2. ENCEPP CHECKLIST FOR STUDY PROTOCOLS

Section 1: Research question	**Yes**	**No**	**N/A**	**Page Number(s)**
1.1 Does the formulation of the research question clearly explain: 1.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue) 1.1.2 The objectives of the study?	☐ ☐	☐ ☐	☐ ☐	
1.2 Does the formulation of the research question specify: 1.2.1 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised) 1.2.2 Which formal hypothesis(-es) is (are) to be tested? 1.2.3 if applicable, that there is no *a priori* hypothesis?	☐ ☐ ☐	☐ ☐ ☐	☐ ☐ ☐	

Comments:

Section 2: Source and study populations	**Yes**	**No**	**N/A**	**Page Number(s)**
2.1 Is the source population described?	☐	☐	☐	
2.2 Is the planned study population defined in terms of: 2.2.1 Study time period? 2.2.2 Age and sex?	☐ ☐	☐ ☐	☐ ☐	

Section 2: Source and study populations	Yes	No	N/A	Page Number(s)
2.2.3 Country of origin?	☐	☐	☐	
2.2.4 Disease/indication?	☐	☐	☐	
2.2.5 Co-morbidity?	☐	☐	☐	
2.2.6 Seasonality?	☐	☐	☐	
2.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☐	☐	☐	

Comments:

Section 3: Study design	Yes	No	N/A	Page Number(s)
3.1 Does the protocol specify the primary and secondary (if applicable) endpoint(s) to be investigated?	☐	☐	☐	
3.2 Is the study design described? (e.g. cohort, case-control, randomised controlled trial, new or alternative design)	☐	☐	☐	
3.3 Does the protocol describe the measure(s) of effect? (e.g. relative risk, odds ratio, deaths per 1000 person-years, absolute risk, excess risk, incidence rate ratio, hazard ratio, number needed to harm (NNH) per year)	☐	☐	☐	
3.4 Is sample size considered?	☐	☐	☐	
3.5 Is statistical power calculated?	☐	☐	☐	

Comments:

Section 4: Data sources	Yes	No	N/A	Page Number(s)
4.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
4.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview, etc)				
4.1.2 Endpoints? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics, etc)	☐	☐	☐	
4.1.3 Covariates?	☐	☐	☐	
	☐	☐	☐	
4.2 Does the protocol describe the information available from the data source(s) on:				
4.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)				
4.2.2 Endpoints? (e.g. date of occurrence, multiple event, severity measures related to event)	☐	☐	☐	
4.2.3 Covariates? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, life style, etc.)	☐	☐	☐	
	☐	☐	☐	
4.3 Is the coding system described for:				
4.3.1 Diseases? (e.g. International Classification of Diseases (ICD)-10)				
4.3.2 Endpoints? (e.g. Medical Dictionary for	☐	☐	☐	

Section 4: Data sources	Yes	No	N/A	Page Number(s)
Regulatory Activities(MedDRA) for adverse events) 4.3.3 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC)Classification System)	☐ ☐	☐ ☐	☐ ☐	
4.4 Is the linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	☐	

Comments:

Section 5: Exposure definition and measurement	Yes	No	N/A	Page Number(s)
5.1 Does the protocol describe how exposure is defined and measured? (e.g. operational details for defining and categorising exposure)	☐	☐	☐	
5.2 Does the protocol discuss the validity of exposure measurement? (e.g. precision, accuracy, prospective ascertainment, exposure information recorded before the outcome occurred, use of validation sub-study)	☐	☐	☐	
5.3 Is exposure classified according to time windows? (e.g. current user, former user, non-use)	☐	☐	☐	
5.4 Is exposure classified based on biological mechanism of action?	☐	☐	☐	
5.5 Does the protocol specify whether a dose-dependent or duration-dependent response is measured?	☐	☐	☐	

Comments:

Section 6: Endpoint definition and measurement	Yes	No	N/A	Page Number(s)
6.1 Does the protocol describe how the endpoints are defined and measured?	☐	☐	☐	
6.2 Does the protocol discuss the validity of endpoint measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, prospective or retrospective ascertainment, use of validation sub-study)	☐	☐	☐	

Comments:

Section 7: Biases and Effect modifiers	Yes	No	N/A	Page Number(s)
7.1 Does the protocol address: 7.1.1 Selection biases? 7.1.2 Information biases? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods)	☐ ☐	☐ ☐	☐ ☐	
7.2 Does the protocol address known confounders? (e.g. collection of data on known confounders, methods of controlling for known confounders)	☐	☐	☐	
7.3 Does the protocol address known effect modifiers? (e.g. collection of data on known effect modifiers, anticipated direction of effect)	☐	☐	☐	
7.4 Does the protocol address other limitations?	☐	☐	☐	

Comments:

Section 8: Analysis plan	Yes	No	N/A	Page Number(s)
8.1 Does the plan include measurement of absolute effects?	☐	☐	☐	
8.2 Is the choice of statistical techniques described?	☐	☐	☐	
8.3 Are descriptive analyses included?	☐	☐	☐	
8.4 Are stratified analyses included?	☐	☐	☐	
8.5 Does the plan describe the methods for identifying: 8.5.1 Confounders? 8.5.2 Effect modifiers?	☐ ☐	☐ ☐	☐ ☐	
8.6 Does the plan describe how the analysis will address: 8.6.1 Confounding? 8.6.2 Effect modification?	☐ ☐	☐ ☐	☐ ☐	

Comments:

Section 9: Quality assurance, feasibility and reporting	Yes	No	N/A	Page Number(s)
9.1 Does the protocol provide information on data storage? (e.g. software and IT environment, database maintenance and anti-fraud protection, archiving)	☐	☐	☐	
9.2 Are methods of quality assurance described?	☐	☐	☐	
9.3 Does the protocol describe quality issues related	☐	☐	☐	

Section 9: Quality assurance, feasibility and reporting	**Yes**	**No**	**N/A**	**Page Number(s)**
to the data source(s)?				
9.4 Does the protocol discuss study feasibility? (e.g. sample size, anticipated exposure, duration of follow-up in a cohort study, patient recruitment)	☐	☐	☐	
9.5 Does the protocol specify timelines for				
9.5.1 Start of data collection?	☐	☐	☐	
9.5.2 Any progress report?	☐	☐	☐	
9.5.3 End of data collection?	☐	☐	☐	
9.5.4 Reporting? (i.e. interim reports, final study report)	☐	☐	☐	
9.6 Does the protocol include a section to document future amendments and deviations?	☐	☐	☐	
9.7 Are communication methods to disseminate results described?	☐	☐	☐	
9.8 Is there a system in place for independent review of study results?	☐	☐	☐	

Comments:

Section 10: Ethical issues	**Yes**	**No**	**N/A**	**Page Number(s)**
10.1 Have requirements of Ethics Committee/Institutional Review Board approval been described?	☐	☐	☐	
10.2 Has any outcome of an ethical review procedure been addressed?	☐	☐	☐	
10.3 Have data protection requirements been described?	☐	☐	☐	

Comments:
Name of main author of study protocol: _____
Date: / /
Signature: _____