BMJ open accessible medical research

Seasonal Influenza Vaccine Effectiveness in the community (SIVE): protocol for a cohort study exploiting a unique national linked data set

Nazir I Lone,[1] Colin Simpson,[1] Kimberley Kavanagh,[2] Chris Robertson,[2,3] Jim McMenamin,[3] Lewis Ritchie,[4] Aziz Sheikh[1]

To cite: Lone NI, Simpson C, Kavanagh K, *et al*. Seasonal Influenza Vaccine Effectiveness in the community (SIVE): protocol for a cohort study exploiting a unique national linked data set. *BMJ Open* 2012;**2**: e001019. doi:10.1136/bmjopen-2012-001019

► Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2012-001019).

Received 10 February 2012
Accepted 17 February 2012

[1]Centre for Population Health Sciences, The University of Edinburgh, Edinburgh, UK
[2]Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK
[3]Health Protection Scotland, Glasgow, UK
[4]Centre of Academic Primary Care, University of Aberdeen, Aberdeen, UK

Correspondence to
Dr Nazir I Lone;
nazir.lone@ed.ac.uk

ABSTRACT

Introduction: Seasonal influenza vaccination is recommended for all individuals aged 65 years and over and in individuals younger than 65 years with comorbidities. There is good evidence of vaccine effectiveness (VE) in young healthy individuals but less robust evidence for effectiveness in the populations targeted for influenza vaccination. Undertaking a randomised controlled trial to assess VE is now impractical due to the presence of national vaccination programmes. Quasi-experimental designs offer the potential to advance the evidence base in such scenarios, and the authors have therefore been commissioned to undertake a naturalistic national evaluation of seasonal influenza VE by using data derived from linkage of a number of Scottish health databases. The aim of this study is to examine the effectiveness of the seasonal influenza vaccination in the Scottish population.

Methods and analysis: A cohort study design will be used pooling data over nine seasons. A primary care database covering 4% of the Scottish population for the period 2000–2009 has been linked to the national database of hospital admissions and the death register and is being linked to the Health Protection Scotland virology database. The primary outcome is VE measured in terms of rate of hospital admissions due to respiratory illness. Multivariable regression will be used to produce estimates of VE adjusted for confounders. The major challenge of this approach is addressing the strong effect of confounding due to vaccinated individuals being systematically different from unvaccinated individuals. Analyses using propensity scores and instrumental variables will be undertaken, and the effect of an unknown confounder will be modelled in a sensitivity analysis to assess the robustness of the estimates.

Ethics and dissemination: The West of Scotland Research Ethics Committee has classified this project as surveillance. The study findings will be disseminated in peer-reviewed publications and presented at international conferences.

ARTICLE SUMMARY

Article focus

- Study protocol for a cohort study to investigate the effectiveness of the seasonal influenza vaccine in the general population.

Key messages

- Seasonal influenza is responsible for substantial global morbidity and mortality, particularly in high-risk populations. Uptake rates for seasonal influenza vaccine remain suboptimal.
- As randomised controlled trials are no longer feasible to assess VE, quasi-experimental methods can be used in their place.

Strengths and limitations of this study

- The study population comprises a large unbiased sample of the general population.
- We are developing a unique linked national database, which contains anonymised individual patient-level data from general practices, hospitals, virology investigations and the death register.
- Our analysis plan takes a robust and comprehensive approach to the well-described problem of confounding in VE studies.
- As this is an observational study, residual confounding may still be present despite the comprehensive approach we plan to take to deal with this.

INTRODUCTION

Each year, influenza causes substantial morbidity and mortality, particularly in people aged 65 years and over and those with underlying serious comorbidities. In the USA, it has been estimated that influenza is responsible for 186 000 excess hospitalisations and 44 000 excess deaths.[1] National vaccination strategies represent a potentially important approach to reduce both influenza-related illness and death, hence the considerable investment in this approach in many parts of the world. Although

vaccination rates in those over 65 in Scotland are a reasonable 75.0% (season 2009/2010), the rates in at-risk groups younger than 65 years remain low (53.4% in season 2009/2010) despite widely promulgated guidelines and incentivised vaccination programmes.[2] There is good evidence of the benefits of the vaccine in young healthy adults and children,[3] but a scarcity of reliable estimates from randomised controlled trials in at-risk populations.[4] There is also limited evidence from observational research, which has only shown effectiveness of vaccination in selected groups of patients, for example, those aged over 65 years[5] or those in at-risk groups for single influenza seasons.[6] Furthermore, these studies may have been prone to bias and residual confounding.[4 7] This may explain, in part, the reason for lower vaccine uptake rates.

Randomised controlled trials offer the best opportunity to produce unbiased estimates of vaccine effectiveness (VE). However, given that influenza vaccination programmes exist in most developed countries, this form of study design is now impractical and is viewed by many in the medical community as unethical.[8] Observational studies are an alternative to investigate VE. However, an individual's decision to attend the local general practice surgery for vaccination may be a marker of healthier behaviour generally, as well as identifying more highly educated individuals who are more aware of and more likely to act on recommendations for their own health. These individuals may be less likely to die from any cause or be admitted to hospital, thus inducing a spurious relationship between vaccination status and the outcome (ie, positive confounding). Similarly, patients who are very frail and unable to attend the general practice surgery may be less likely to be vaccinated, but much more likely to die or be admitted to hospital.[9] This phenomenon is also known as the 'healthy vaccine effect'.

Standard methods of adjustment for confounders are likely to be inadequate to control for confounding due to the healthy vaccine effect. This can result in excessive estimates of VE in observational studies using non-influenza-specific outcomes due to residual confounding. A number of methods can be used to try and address this problem, including quasi-experimental study designs and advanced statistical methods. In addition, an analysis framework has been proposed to identify residual confounding when undertaking VE studies using observational methods.[10]

This research aims to examine the effectiveness of the seasonal influenza vaccine while addressing the methodological challenges outlined above of using observational data. We will have access to a unique set of linked databases, which contain individual patient-level data relating to primary healthcare, acute hospital care, virological laboratory tests and mortality. In contrast to previous observational studies, these rich data sources provide information on a large number of potential confounders and highly specific laboratory outcome measures in a study cohort sampled from the general population. Our assessment of the effectiveness and impact of the seasonal influenza vaccination programme therefore offers potentially large societal benefits both for Scotland, the UK, and for advancing the international evidence base.

AIMS AND OBJECTIVES

We aim to examine the effectiveness of the seasonal influenza vaccination in individuals registered with a national sample of general practices in Scotland. More specifically, the objectives of this study are to (1) report vaccine uptake in the relevant at-risk populations for whom vaccination is recommended in the UK; (2) evaluate VE measured in terms of the following outcomes: rate of hospital admissions due to respiratory illness (primary outcome), rate of primary care consultations due to respiratory illness, risk of death due to respiratory illness and risk of laboratory-confirmed influenza infection and (3) assess the degree of and adjust for residual confounding in our estimates using analyses incorporating propensity scores, instrumental variables and the effect of a hypothetical unknown confounder.

METHODS

Study design and population

A cohort study design will be used to assess VE. Vaccine uptake will be reported using serial cross-sectional surveys. Data extracted from 35 general practices of the sentinel surveillance network in Scotland, the Practice Team Information network, will be used. Participating practices cover a 4% sample of the Scottish population (n=209 452 registered alive in 2009). The population targeted for influenza vaccination comprises all patients aged 65 years and older (approximately 15% of the general population, n=28 241 in the sample) and those aged younger than 65 years defined as being in an at-risk group on the basis of pre-existing illness (n=~33 000, 18% of younger than 65-year-olds).[11] The estimates for proportion of patients younger than 65 years in an at-risk group were taken from our recent VIPER study investigating the effectiveness of the 2009 H1N1 pandemic influenza vaccine.[12] Each patient will contribute person-time to each influenza season while alive and registered with a participating general practice. The primary care database was linked to the Scottish acute hospital discharge database and Scottish death register as part of the VIPER project.[12] In addition, a linkage of these data sets to the Health Protection Scotland virology database to determine laboratory-confirmed influenza infection is underway (due to be completed on 1 February 2012).

Databases

Acute hospital discharge database

The Information Services Division, National Services Scotland, maintains a database of all acute hospital discharges in Scotland, known as the Scottish Morbidity Record 1. All inpatient and day case episodes of care for

acute hospitals since 1981 have been recorded in the database. The database is subject to regular validation checks, and the most recent quality assurance report indicated good levels of accuracy (>90%) for the fields used in this study.[13] Diagnostic information is recorded using International Classification of Disease version 10 (ICD-10). There are up to six fields that can be used to record diagnoses, with one allocated as the main reason for admission. Scottish Morbidity Record 1 is linked routinely by Information Services Division to the Scottish death register using patient characteristics in a probabilistic matching algorithm with a high degree of accuracy.[14 15]

Primary care database

Almost all individuals resident in Scotland are registered with a primary care practice, which provides healthcare services free of charge. Virtually all specialist hospital care services are also free of charge, usually obtained through referral from primary care or, in emergency situations, through patients attending an emergency department. Primary care-based physicians provide or coordinate much of the care of patients discharged back into the community by secondary and tertiary care services. The primary care database was linked to the other databases using probabilistic linkage. Linkage accuracy was high due to the high quality and number of patient identifiers available from the primary care database. Completeness of capture of contacts and accuracy of clinical event coding (using Read codes) has been found to be above 91% among the study practices.[16 17] The electronic recording of long-term prescribing information by primary care has also been found to be both accurate and complete.[18]

Death register

Details from death certificates issued for all deaths in Scotland are recorded in the death register, maintained by National Records Scotland.[19] Cause of death has been routinely coded using ICD-10 since 2000.

Health Protection Scotland virology database

The Scotland-wide unique patient identifier, the Community Health Index number, is being used to link records in the virology database to the other databases. There may be fewer virology records with adequate patient identifiable information to allow linkage for the years 2000–2005 of the study period. This is actively being investigated by coinvestigators.

Study period

Data from 1 September 2000 to 31 August 2009 will be used. This will allow analysis of nine influenza seasons (2000/2001 to 2008/2009). Each year (1 September to 31 August) will be divided into four periods (figure 1). The influenza season will be defined for each year using national influenza surveillance data.[20] The start of the period will be from the date of the first influenza isolate reported by Health Protection Scotland each year. The end of the influenza season will be the date of the last influenza isolate with an additional 14-day period for complications. The pre-influenza period will be defined as starting from 1 September each year until the date of the first influenza isolate. The post-influenza period will start 14 days after the last influenza isolate and end on 31 May each year. The 'non-influenza' period for each year will be from 1 June to 31 August (figure 1).

Exposure definition

Vaccination will be used to define exposure status if it is given at a time point between the start of the pre-influenza season (1 September) and the end of the influenza season (figure 1). An individual will be defined as vaccinated 14 days after the seasonal influenza vaccine has been administered.[21] The time period from the first day of the influenza season to day 14 post-vaccination will be defined as 'unexposed' and the period from day 14 post-vaccination until the end of the influenza season will be defined as 'exposed'. Therefore, those vaccinated between the start of the pre-influenza period up until 14 days before the influenza season will be defined as 'exposed' for the duration of the influenza season.

Outcomes

VE should ideally be measured using influenza-specific outcomes in each of the databases. However, it is likely that ICD codes or Read codes referring to influenza-specific outcomes are underused by clinicians and coders, thereby reducing the sensitivity and power of the primary analysis. For this reason, codes for acute respiratory diseases were chosen as primary outcome measures as they would capture a substantial proportion of

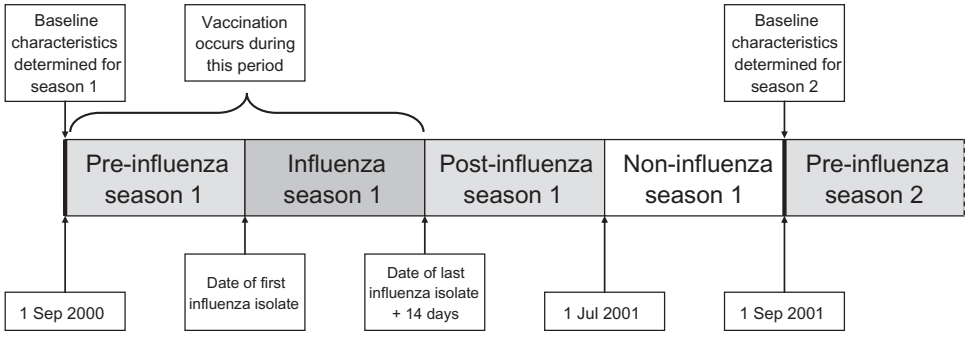

Figure 1 Relationship of first influenza season (2000–2001) to pre-, post- and non-influenza season periods. Baseline characteristics for each patient are determined on 1 September each year. The earliest date of influenza vaccination varied for each influenza season but always occurred after 1 September.

influenza-related events during the influenza season. Laboratory-confirmed influenza infection will be a highly specific outcome and will be calculated on a subgroup of the study population (see Laboratory methods section).

VE will be calculated by subtracting the rate ratio (RR) or odds ratio (OR) of vaccinated compared with unvaccinated patients from 1 (ie, VE=(1−OR)×100% or VE=(1−RR)×100%) for each of the following outcome measures:

a. Hospital discharge data: rate of emergency hospitalisations with a diagnosis of influenza or pneumonia (primary measure of VE).
b. Primary care data: rate of consultations in primary care for influenza-like illnesses and acute respiratory infection.
c. Death register: deaths due to influenza, pneumonia or chronic obstructive pulmonary disease (COPD).
d. Health Protection Scotland Virology Database: laboratory-confirmed influenza.

Additional sensitivity analyses will be undertaken using less-specific outcomes (all-cause mortality, emergency admission to hospital for any reason) as well as the influenza-specific outcomes (deaths due to influenza, hospital admissions due to influenza). These analyses will be part of the framework to assess bias (see below).

A number of secondary analyses will be undertaken using other outcomes. The effect of vaccination status on hospital admissions and deaths relating to cardiovascular and cerebrovascular events as a composite outcome will be analysed. In addition, exploratory analyses will be undertaken to assess the effect of vaccination status on outcomes for which it would not be expected to have an effect, for example, appendicitis or trauma. This approach of using an alternative outcome as a negative control has been shown to be a useful method for detecting residual bias.[22]

Confounding factors

Individuals who are vaccinated are likely to be different from those who are unvaccinated. An extensive list of individual-level and practice-level characteristics will be included as confounders in the analyses. These will be defined in each year on the first day of the pre-influenza season (1 September). Practice effects will be accounted for using multilevel methods.

Demographics

Sex, age-band (0–4, 5–14, 15–44, 45–64, 65–74 and 75+) and socioeconomic status will be included in all analyses; socioeconomic status will be measured using quintiles of the Scottish Index of Multiple Deprivation: 1=most affluent and 5=most deprived). Scottish Index of Multiple Deprivation is an area-based measure of deprivation derived from seven domains including income, employment and education.[23]

At-risk groups

At-risk patients are those with certain comorbidites for whom seasonal influenza vaccination is indicated. Patients will be defined as high risk according to national guidance[11] if they have one or more of the following conditions:

- chronic heart disease,
- chronic kidney disease (including renal transplantation),
- chronic liver disease,
- chronic neurological disease,
- chronic respiratory disease,
- conditions or drugs causing impaired immune function and
- diabetes.

Chronic diseases

Comorbidity will be defined by the 17 disease categories that constitute the Charlson Comorbidity Index.[24] This index has been validated in a number of different databases using codes from healthcare databases.[25] A recent study has mapped Read codes from a UK general practice database to the relevant Charlson comorbid disease groups, resulting in a model that performed well in the prediction of 5-year mortality.[26] These codes will be used to identify comorbidities that are present in a patent's record prior to the start of each pre-influenza season (1 September). The number of repeat prescription items issued in the previous 12-month period will be used as an additional measure of comorbidity. Measures of previous healthcare resource use will also capture other aspects of chronic health status.

Smoking status

This will be derived from primary care data (current smoker, ex-smoker, non-smoker) and determined on 1 September each year.

Previous vaccinations

A variable will be included for patients who have received seasonal influenza vaccination in the previous season to account for the possibility of persisting VE in the subsequent year.[27] Adjustment for previous pneumococcal vaccination at any time in the primary care record prior to 1 September each year will also be undertaken, which will be particularly important for less-specific outcome measures such as all-cause death.

Previous healthcare utilisation

The number of general practice consultations in the previous 12 months will be used as a measure of healthcare utilisation. Number of emergency admissions to hospital for any cause during the previous 12 months will be used as a marker of severity of chronic health status.

Functional status

There is no direct measure of functional status made in any of these national databases. However, individuals who require a home visit when consulting a general practitioner rather than attending the practice will be identified. This will be used as a proxy marker for poor

mobility or frailty. In addition, individuals who are resident in some form of institutional care setting may be identifiable from the primary care database. This will also be used as an indicator of more severe functional limitation.

Laboratory methods

Not all patients receiving vaccination will have been swabbed for influenza. The majority of Practice Team Information general practices are involved in the Health Protection Scotland sentinel swabbing scheme, whereby practices are encouraged to submit five swab samples per week to the West of Scotland Specialist Virology Centre (for multiplex polymerase chain reaction testing for a range of respiratory pathogens) on any patient presenting for consultation in the practice for influenza-like illness and acute respiratory infections across all ages. Crucially, this is independent of whether the individual has or has not been vaccinated.

Subgroups for stratified analyses

The primary analysis will be performed using the whole cohort. A stratified analysis will be undertaken for those aged 65 years and older and those aged younger than 65 years old and by risk groups. This will allow VE to be assessed in more homogeneous subgroups and to check for effect modification across strata. Further subgroup analysis will be performed by restricting the study cohort to the main population group for whom the national influenza vaccination programme is targeted: those aged 65 years or older together with those aged younger than 65 years in an at-risk group.[11] It is likely that these analyses in subgroups will be underpowered, in particular in the younger than 65-year age group for whom event rates will be lower. Further stratified analyses will be undertaken as part of the framework to assess residual confounding (see section below).

Methods to further adjust for confounding

We plan to take a comprehensive approach to dealing with confounding due to the healthy vaccine effect in our study. This will include the use of complex statistical methods including propensity score analysis, instrumental variable analysis and modelling a hypothetical unmeasured confounder. These are considered in more detail below.

Propensity scores

Propensity scores are a well-described method to reduce the effect of strong confounding, such as confounding by indication.[28] We will develop a propensity score to predict the likelihood that a patient receives the seasonal influenza flu vaccine. This will allow for a better comparison to be made between vaccinated and unvaccinated groups. A logistic regression model will be constructed with vaccination status as the outcome in order to produce a score of the propensity to be vaccinated. The covariates in the model will be derived from available patient- and practice-level characteristics, which we consider to be clinically relevant to the probability of receiving the vaccination.

Instrumental variables

Instrumental variable analyses are well established in non-healthcare settings such as econometrics as a means of adjusting for unmeasured confounding.[29] An instrumental variable is a factor related to exposure status (ie, vaccination status), which does not have an independent effect on outcome other than by ways mediated through the exposure. Furthermore, an instrumental variable should not be related to any variables that confound the relationship between exposure and outcome. There are no instrumental variables that have been established as appropriate for analyses of VE.[30] For this reason, we will explore a number of variables that may fulfil the above criteria on a conceptual level.

Sensitivity analysis for unmeasured confounding

We will carry out a sensitivity analysis on the results of our primary analysis to explore the impact of an unmeasured confounding factor on the estimate of VE.[31] This will allow us to account for confounders such as poor functional status which is, as noted above, incompletely recorded in the national databases. This method assumes that the unmeasured confounder is not associated with the measured confounders in the model and is therefore likely to overestimate the impact of the unmeasured confounder.[32]

Framework for detecting residual confounding

As described earlier, we will undertake additional analyses to identify the presence of residual confounding. This has been recommended as part of an analytical framework when reporting VE using observational study designs.[33] We will assess the variation in VE using the following criteria:

1. Seasonality: stratification on season is more important when VE is measured using non-specific outcomes. Each year of observation will be partitioned into four periods: non-influenza period, pre-influenza season, influenza season (when influenza virus is circulating) and post-influenza season (figure 1). Maximal VE should be seen during the influenza season. The vaccine should have no effect on outcome in the pre-influenza and non-influenza seasons. The non-influenza season will use vaccination status from the previous influenza season. This is to minimise the bias that might occur when vaccination status is applied retrospectively. This retrospective application of vaccine status would include patients who die during the preceding non-influenza season as unvaccinated, despite the fact that they would not have survived long enough to be eligible for vaccination.
2. Vaccine match: VE should be lower in years during which the influenza vaccine was a poor match for the circulating virus.

3. Severity of influenza season: VE should be greater in years during which the circulating virus caused a large excess mortality during the influenza season.
4. Age: it is thought that influenza vaccine is less effective in the oldest age groups due to immune senescence.[33] If this assumption is correct, VE should be lowest in the oldest subgroup. A stratified analysis on age groups will be undertaken to assess for this effect.
5. Specificity of outcome measure: VE should be greatest for the most specific outcome (laboratory-confirmed influenza infection) and lowest for the less-specific outcomes (all-cause mortality). In addition to the primary analysis, in the three non-laboratory databases, we will undertake analyses using the more influenza-specific outcomes (influenza-coded deaths, hospital admissions and primary care attendances) and less-specific outcomes (all-cause deaths, any emergency hospital admissions and any primary care attendances).

Statistical analysis

Baseline characteristics will be summarised by vaccination status for the whole cohort using mean, median or proportion where appropriate together with a measure of dispersion. Missing data will be reported for each variable. A 5% significance level will be used for hypothesis tests for the primary outcome. All p values will be two sided. All analyses will be undertaken in R. CIs for the RR and tests of the differences between two rates will be carried out using the 'midp method' in the RR function and rate2by2.test function, respectively, using the 'epitools' package in R.[34] For small samples, CIs for the RR will be estimated using the Excel workbook. Effect modification will be assessed across age groups, year and season by entering an interaction term into models. A complete case analysis will be the primary analysis. Multiple imputation using chained equations will be used, if necessary, to perform analyses on imputed data sets to assess the effect of missing data on VE estimates.[35]

Annual and pooled analyses

We will initially analyse each of the nine influenza seasons from 2000/2001 to 2008/2009 separately to calculate the VE for each season. We plan to test the homogeneity of the vaccine effect over the seasons, and if appropriate, pool the data to give a more powerful analysis than would be obtained using a simple aggregation of data. In the pooled analysis, we will account for the within-person correlation resulting from repeated measures on the same individual in subsequent seasons by the use of generalised estimation equations or an adjustment for clustering. This is likely to be more computationally efficient than hierarchical models. For the pooled analysis, we will be able to incorporate the effects of time and also to let the vaccine effect vary with year yet keeping the effect of the explanatory variables constant over time. If a pooled analysis is not considered appropriate, for example, if there is evidence of substantial heterogeneity in the VE over seasons, then we will use meta-regression models to try to explain this heterogeneity.

Vaccine uptake

ORs (adjusted for age, sex and deprivation) will be calculated for differences in vaccine uptake rates between different groups of patients (sex, age, deprivation quintiles and at-risk groups).

Vaccine effectiveness

Crude and adjusted VE estimates will be reported for each outcome. VE estimates will be calculated for the cohort as a whole and stratified using the subgroups specified above. We will use the VE outcomes above to calculate numbers needed to vaccinate to prevent one swab-determined influenza infection, hospitalisation, consultation and death. A person-time denominator will be used for general practice consultations, hospital admissions and death. Follow-up time will be censored at death from any cause for consultations and admissions. Hospital admissions and consultations can have multiple events and each event will be counted.

Hospitalisations and primary care consultations

The ratio of the number of admissions to hospital per person-time during the post-vaccination period compared with the number of admission to hospital per person-time during the pre-vaccination period will be calculated. The unadjusted estimate of VE will be calculated as $(1-\text{RR})\times 100\%$. Adjusted RRs of VE for prevention of hospitalisation will be derived from Poisson regression models, adjusting for the confounders listed above. Similar methods will be used to estimate VE for primary care consultations.

Deaths

The OR of deaths in the vaccinated group to deaths in the unvaccinated group will be calculated; these will be both unadjusted and adjusted for the confounders listed above. VE will be calculated as $(1-\text{OR})\times 100\%$.

Laboratory-confirmed infection

For VE, using information from linked virological swab data, a logistic regression model will be fitted adjusting for the confounders listed above. VE will be measured by comparing swabs taken after vaccination with swabs taken before vaccination for all vaccinated individuals and second by comparing swabs taken after vaccination among those vaccinated to swabs taken among those never vaccinated. VE will be calculated as $(1-\text{OR})\times 100$.

Statistical methods to further adjust for confounding

Propensity score

We will undertake analyses incorporating propensity scores using three different methods: regression (including propensity score as a covariate), stratification (based on quintile of propensity score) and matching (vaccinated and non-vaccinated patients individually matched by propensity score). The model will be

non-parsimonious in order to include a wide range of factors that influence propensity to be vaccinated. The following covariates will be included in the model: age, sex, socioeconomic status, comorbidities for which vaccination is indicated (see above), comorbidities included in the Charlson Index, smoking status, previous vaccinations, functional status/frailty, number of primary care consultations in previous year, number of hospitalisations in previous year, number of repeat prescription items in previous year, practice type, overall practice deprivation, practice vaccination rate and influenza season.

Instrumental variable analysis

We will consider the following variables as potential instrumental variables: previous antacid prescription, previous thyroxine prescription, gout and screening attendance. We will assess whether these variables fulfil the following criteria for use as an instrumental variable: association with vaccination status (exposure), no association with outcome other than thorough exposure and no association with confounding variables. The rationale for selecting these variables is that each may increase the likelihood of a patient being opportunistically vaccinated while attending the general practitioner but should not be related to the risk of contracting influenza. If a suitable instrumental variable is found, analyses will be undertaken to produce VE estimates adjusted for the measured confounders and the instrumental variable. We will use a two-stage estimation method with logistic models used to combine the two equations.

Modelling an unmeasured confounder

Death rates and hospital admission rates are likely to be highest in the frailest members of the study population. As these patients are less likely to seek vaccination, it has been suggested that inadequately measured frailty may explain some of the VE measured in observational studies.[7 9] As we may have been unable to fully account for frailty, which has been defined in recent studies,[36 37] we will use estimates from published data to model this unmeasured (or inadequately measured) confounder in a sensitivity analysis. We will assume that prevalence of frailty varies from 5% to 20% in those aged 65 years or older,[36 37] that frail individuals are two to four times more likely to be hospitalised or die[5] and assume that frail individuals have a 50% lower probability of being vaccinated.[38]

Sample size

In our related VIPER study,[12] the VE estimates were of the order 50% or greater, depending upon the end point, even though the follow-up time was limited to 90 days after vaccination. The power calculations are based upon a comparison of rates, assuming a Poisson distribution, for consultations, hospitalisations and deaths. Baseline rates are derived from previous studies for hospitalisations and consultations in VIPER[12] and surveillance data from the Pandemic Influenza Primary Care Reporting system.[39] For mortality, rates are derived from statistics published by National Records Scotland.[40] For the virological response, the power is derived from the comparison of two proportions and baseline swab positivity derived from Hardelid *et al*.[41] In all cases, an adjustment for the effective population size is made using design effects, estimated from Pandemic Influenza Primary Care Reporting and the VIPER study, ranging from 1.07 to 1.15 associated with the clustering of patients within general practices. Power is calculated for a single year and also for the whole 9-year period, assuming that the vaccine effect is similar in all seasons. For hospitalisations, deaths and virology we anticipate using a 6-month comparison period within each season, while for general practice consultations, a 1-month period is used in the power calculation, although in the analysis a longer time will be used. Power calculations are summarised in table 1.

Table 1 Summary of power calculations for each outcome measure

			One season		Over 9 seasons	
	Rate/100 000	**% Vaccinated**	**VE**	**Power**	**VE**	**Power**
Hospitalisatons						
Influenza	50	15	70	83	30	79
Influenza and pneumonia	275	15	40	86	15	85
Mortality						
Respiratory only (65+ only)	750	60	50	86	20	92
All causes (65+ only)	5000	60	20	83	7	85
Primary care consultations						
ILI consultations (rate/week)	30	15	40	84	15	82
	Percent positive	**% Vaccinated**	**VE**	**Power**	**VE**	**Power**
Virology						
800 swabs per annum	30	15	45	87	15	83
400 swabs per annum	30	15	60	85	20	79

Power calculations are presented for one season and for pooling across nine seasons. Allowance has been made for the effect of clustering of patients within general practices.
ILI, influenza-like illness; VE, vaccine effectiveness.

Hospitalisations due to respiratory disease

Hospitalisations for influenza are rare at 50/100 000/year and for a single year we will have a power of 83% to detect a 70% vaccine effect. Combining the 9 years of data gives a power of 80% for a 30% vaccine effect. By aggregating both influenza and pneumonia hospital admission, the underlying rate increases over fivefold, and the power to detect a vaccine effect of 40% in a single year is 86% and 85% for a 15% vaccine effect over 9 years (table 1).

Mortality

All cause

Among people over 65 years, the death rate from all causes in Scotland is 5000/100 000. Assuming that 55%–75% of the age group are vaccinated, then this study has a power in excess of 80% to detect a difference of 20% or more in the proportions of vaccinated and unvaccinated individuals dying over a period of 6 months. Assuming similar vaccine effect in each of the 9 years, the power approaches 90% to detect a 7% vaccine effect.

Respiratory deaths

Reductions in respiratory deaths will also be included as a compound outcome. These account for 15% deaths. The estimated 30% reduction in mortality between vaccinated than unvaccinated individuals is a conservative estimate based on data from previous research (eg, Nichol *et al*[5] found a 58% reduction in mortality between vaccinated and unvaccinated in over 65s). For the compound respiratory deaths end point, we will have at least 80% power to detect a 50% mortality reduction in a single year and over 90% power to detect a 20% vaccine effect over the 9-year period.

Primary care consultations for influenza-like illness and acute respiratory infection

Consultation rates for influenza-like illnesses and acute respiratory infections are of the order of 30/100 000/day and over the period of 1 month the whole cohort will have a power of 84% to detect a difference of 40% in consultation rates between the vaccinated and unvaccinated. Extending to nine seasons, the power is over 80% to detect a 15% vaccine effect. Extending the follow-up time each season will increase the power. Vaccine uptake is assumed to be at 15% of the whole population.

Laboratory-confirmed influenza infection

During the period of influenza activity, swab positivity for influenza is around 30%. If there are 800 swabs collected per year and there is a 15% vaccine uptake, then a vaccine effect of 45% can be detected in a single season with a power of 87% and a vaccine effect of 15% detected over 9 years with a power of 83%. With fewer swabs taken, the power is smaller though even if only 400 swabs are taken per year, a vaccine effect of 20% can be detected with just under 80% power.

ETHICS AND DISSEMINATION

The West of Scotland Research Ethics Committee has classified this project as surveillance. The Privacy Advisory Committee of the Information Services Division, National Services Scotland, approved the linking of the anonymised data sets. Each of the 35 general practices gave consent for the extraction and use of primary care data. An Independent Steering Committee has been convened to oversee this research. The study findings will be disseminated in peer-reviewed publications and presented at international conferences.

Contributors NIL, CS, CR, JM, LR and AS contributed to the conception of the study. All authors contributed to the study design. All authors contributed to drafting the protocol. All authors revised the manuscript for important intellectual content. All authors gave final approval of the version to be published.

Funding The study is funded by a project grant from the National Institute for Health Research Health Services Research programme (09/2000/37). The views expressed in this manuscript are those of the authors and not necessarily those of the NIHR.

Competing interests None.

Ethics approval The West of Scotland Research Ethics Committee classified the project as surveillance and therefore waived the need for ethical approval.

Provenance and peer review Not commissioned; internally peer reviewed.

REFERENCES

1. Thompson WW, Shay DK, Weintraub E, *et al*. Mortality associated with influenza and respiratory syncytial virus in the United States. *JAMA* 2003;289:179–86.
2. Information Services Division. *Influenza Vaccinations: Vaccination Uptake*. National Services Scotland, 2010. http://www.isdscotlandarchive.scot.nhs.uk/isd/6365.html (accessed 3 Sep 2011).
3. Smith S, Demicheli V, Di Pietrantonj C, *et al*. Vaccines for preventing influenza in healthy children. *Cochrane Database Syst Rev* 2006;(1):CD004879.
4. Jefferson T, Di Pietrantonj C, Al-Ansary L, *et al*. Vaccines for preventing influenza in the elderly. *Cochrane Database Syst Rev* 2010;(2):CD004876.
5. Nichol KL, Nordin JD, Nelson DB, *et al*. Effectiveness of influenza vaccine in the community-dwelling elderly. *N Engl J Med* 2007;357:1373–81.
6. Hak E, Buskens E, Nichol KL, *et al*. Do recommended high-risk adults benefit from a first influenza vaccination? *Vaccine* 2006;24:2799–802.
7. Jackson LA, Jackson ML, Nelson JC, *et al*. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. *Int J Epidemiol* 2006;35:337–44.
8. Hak E, Verheij TJ, Grobbee DE, *et al*. Confounding by indication in non-experimental evaluation of vaccine effectiveness: the example of prevention of influenza complications. *J Epidemiol Community Health* 2002;56:951–5.
9. Simonsen L, Viboud C, Taylor RJ, *et al*. Influenza vaccination and mortality benefits: new insights, new opportunities. *Vaccine* 2009;27:6300–4.
10. Simonsen L. Commentary: observational studies and the art of accurately measuring influenza vaccine benefits. *Int J Epidemiol* 2007;36:631–2.
11. Chief Medical Officer. *Seasonal Influenza Vaccination Programme 2010–11*. Reference:SGHD/CMO, (2010)12. Scottish Government, 2010. http://www.sehd.scot.nhs.uk/cmo/CMO(2010)14.pdf
12. Simpson CR, Ritchie LD, Robertson C, *et al*. Vaccine effectiveness in pandemic influenza - primary care reporting (VIPER): an observational study to assess the effectiveness of the pandemic influenza A (H1N1)v vaccine. *Health Technol Assess* 2010;14:313–46.
13. Information Services Division. *NHS Hospital Data Quality: Towards Better Data From Scottish Hospitals*. Edinburgh: ISD Publications Scotland, 2007.
14. Kendrick S, Clarke J. The Scottish record linkage system. *Health Bull* 1993;51:72–9.

15. Scottish Public Health Observatory. *Overview of Key Data Sources: ISD Linked Database*. 2010. http://www.scotpho.org.uk/home/resources/OverviewofKeyDataSources/Nationaldataschemes/Linked_database.asp
16. Information Services Division. *Practice Team Information (PTI) Statistics*. National Services Scotland, 2010. http://www.isdscotland.org/Health-Topics/General-Practice/PTI-Statistics (accessed 12 Aug 2011).
17. McAlister FA, Murphy NF, Simpson CR, *et al*. Influence of socioeconomic deprivation on the primary care burden and treatment of patients with a diagnosis of heart failure in general practice in Scotland: population based study. *BMJ* 2004;328:1110.
18. Whitelaw FG, Nevin SL, Milne RM, *et al*. Completeness and accuracy of morbidity and repeat prescribing records held on general practice computers in Scotland. *Br J Gen Pract* 1996;46:181–6.
19. Scottish Public Health Observatory. *Overview of Key Data Sources: Deaths*. 2010. http://www.scotpho.org.uk/home/resources/OverviewofKeyDataSources/Nationaldataschemes/deaths.asp
20. Health Protection Scotland. *Influenza Seasons*. 2011. http://www.hps.scot.nhs.uk/resp/influenzaseason.aspx (accessed 15 Sep 2011).
21. Influenza (updated November 2011). In: Salisbury D, Ramsay M, Noakes K, eds. *Immunisation against Infections Disease*. London: The Stationery Office, 2006:185–204. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_131001.pdf (accessed 6 Dec 2011).
22. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010;21:383–8.
23. National Statistics. *Scottish Index of Multiple Deprivation: 2009 General Report*. Edinburgh: Scottish Government National Statistics Publications, 2009.
24. Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
25. Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. *Int J Epidemiol* 2000;29:891–8.
26. Khan NF, Perera R, Harper S, *et al*. Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. *BMC Fam Pract* 2010;11:1.
27. Voordouw AC, Sturkenboom MC, Dieleman JP, *et al*. Annual revaccination against influenza and mortality risk in community-dwelling elderly persons. *JAMA* 2004;292:2089–95.
28. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
29. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000;29:722–9.
30. Groenwold RH, Hak E, Klungel OH, *et al*. Instrumental variables in influenza vaccination studies: mission impossible?! *Value Health* 2010;13:132–7.
31. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 1998;54:948–63.
32. Groenwold RH, Nelson DB, Nichol KL, *et al*. Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research. *Int J Epidemiol* 2010;39:107–17.
33. Simonsen L, Taylor RJ, Viboud C, *et al*. Mortality benefits of influenza vaccination in elderly people: an ongoing controversy. *Lancet Infect Dis* 2007;7:658–66.
34. Newcombe R. *Confidence Intervals for Proportions and Related Quantities*. 2010. http://www.cardiff.ac.uk/medic/aboutus/departments/primarycareandpublichealth/ourresearch/resources/index.html (accessed Jan 2010).
35. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;30:377–99.
36. Syddall H, Roberts HC, Evandrou M, *et al*. Prevalence and correlates of frailty among community-dwelling older men and women: findings from the Hertfordshire Cohort Study. *Age Ageing* 2010;39:197–203.
37. Santos-Eggimann B, Cuénoud P, Spagnoli J, *et al*. Prevalence of frailty in middle-aged and older community-dwelling Europeans living in 10 countries. *J Gerontol A Biol Sci Med Sci* 2009;64:675–81.
38. Jackson ML, Nelson JC, Weiss NS, *et al*. Influenza vaccination and risk of community-acquired pneumonia in immunocompetent elderly people: a population-based, nested case-control study. *Lancet* 2008;372:398–405.
39. Information Services Division. *Influenza*. 2010. http://www.isdscotlandarchive.scot.nhs.uk/isd/6365.html (accessed 3 Sep 2011).
40. National Records Scotland. *Weekly Death Registrations—In Total, and Caused by Respiratory System Diseases*. 2011. http://www.gro-scotland.gov.uk/statistics/theme/vital-events/general/weekly-monthly-births-death-data/weekly/weekly-death-registrations-in-total-and-caused-by-respiratory-diseases.html (accessed 1 Oct 2011).
41. Hardelid P, Fleming D, McMenamin J, *et al*. Effectiveness of pandemic and seasonal influenza vaccine in preventing pandemic influenza A(H1N1)2009 infection in England and Scotland 2009-2010. *Euro Surveill* 2011;16:pii: 19763.