Appendix 2 (as supplied by the authors): Supplemental Methods

DATA SOURCES

For this study, we analysed data from three sources linked by the Clinical Practice Research Datalink (CPRD-GOLD; www.cprd.com): the CPRD, the Hospital Episodes Statistics (HES; http://www.hscic.gov.uk/hes), and cause-specific mortality from the Office of National Statistics (ONS; http://www.ons.gov.uk/ons). CPRD provides all health related information recorded and coded in primary care, including demographic and health behaviour data, symptoms, clinical diagnoses, clinical and laboratory tests, medical procedures and prescriptions. CPRD data were coded with the 2nd version of the Read classification system, a hierarchical clinical system containing over 96,00 codes, and data accuracy and completeness are regularly audited. The patients included in CPRD have been shown to be broadly representative of the UK population, in terms of age, sex and ethnicity, when compared with the UK census[1,2]. Patients have been shown to also be comparable in terms of distribution of body mass index, when compared to the Health Survey for England (household based population survey). Validation studies for use of the data in epidemiological research have primarily focused on the study of positive predicted values, which have been reported to be high and comparisons of incidence with other UK data sources have also shown to be similar[3-6]. HES provides information about medical diagnoses during all elective and emergency hospital admission across all National Health Service hospitals in England. Both HES and ONS data are coded with the 10th revision of the International Classification of Diseases (ICD-10). ONS data is coded with the 9th revision of the ICD before the year 2000.

DEFINITIONS OF POLYMYALGIA RHEUMATICA AND GIANT CELL ARTERITIS

Prior studies using primary care data have used a similar approach to ours to defining polymyalgia rheumatic (PMR) and giant cell arteritis (GCA) and used the same diagnostic codes[1,2,3]. The study of Smeeth et al. (2010)[2] showed that a recorded diagnosis of giant cell arteritis in CPRD was supported in 91% of cases (41/45). It is also important to note that more than 60% of patients included in our study had evidence of supporting information for diagnosis of PMR and/or GCA (e.g. a diagnosis in both primary care and during hospitalisation, care and/or referral to a rheumatologist or a temporal biopsy performed to confirm the diagnosis of GCA).

DEFINITION OF COVARIATES

CPRD contains complete information on all prescriptions issued to all the patients registered in the family practices. We defined baseline covariates as follows:

- medication use from CPRD: ≥1 prescriptions issued to the patient within 1 year prior to the start of follow-up
- presence of comorbidities from CPRD and HES: a diagnosis recorded at any time prior to the start of follow-up
- number of hospital visits: number of hospital admissions within 1 year prior to the start of follow-up
- quantitative biomarkers (e.g. body mass index) from CPRD: the closest measurement value recorded within 1 year prior to the start of follow-up
- smoking status from CPRD: the closest status recorded within 1 year prior to the start of follow-up. We then replaced non-smoker status by ex-smoker if current smoker had been recorded for the patient at any time prior to the follow-up start.
- vaccination status from CPRD: vaccination status recorded within 1 year prior to the start of follow-up for influenza; and at any time prior to follow-up start for varicella zoster and pneumococcus

CONSTRUCTION OF TIME-VARIANT MEDICATION VARIABLES

To construct the time-variant variables included in the models, we first identified all prescriptions of each type of medication issued to each patient. Then:

For non-steroidal anti-inflammatory drugs and disease-modifying antirheumatic drugs: For each patient, we determined the start (first prescription during the period) and end of medication periods (last prescription + its duration; with last prescription defined as a 90 day gap without prescription). We then split the study follow-up for each patient on the dates of start/end of each treatment period (so we created a row of data for each period) and created a binary (yes/no) variable to identify whether in each period the patient was or not on medication. This is the standard method to construct time-variant variables.

For oral glucocorticoids, we followed a similar procedure but the split took place on the dates on which the daily dose changed (as indicated in Figure 1).

Appendix to: Wu J, Keeley A, Mallen C, et al. Incidence of infections associated with oral glucocorticoid dose in people diagnosed with polymyalgia rheumatica or giant cell arteritis: a cohort study in England. *CMAJ* 2019. doi: 10.1503/cmaj.190178.

Finally, to integrate time-variant information for the variables for each medication prescribed, when the start and end of periods for different types of medication did not match, we further split the follow period of the patient on the date on which a period started or ended for any of the other medications. In this manner, we ensured that for all patients and for each duration of follow-up, the exposure variables correctly identified whether each patient was receiving or not each type of time-variant medication. A screenshot of the data for selected medication time-variant variables is shown below.

PATID	Age	Sex	tstart	tstop	GC_ever	GC_current	GC_current_dose	DMARD	Infection status
2630	78	Female	0	12	Yes	No	0.0	No	0
2630	78	Female	12	92	Yes	Yes	2.5	No	0
2630	78	Female	92	113	Yes	Yes	4.0	No	0
2630	78	Female	113	334	Yes	No	0.0	No	0
3483	80	Male	0	76	Yes	No	0.0	No	0
3483	80	Male	76	90	Yes	Yes	10.0	No	0
3483	80	Male	90	104	Yes	Yes	2.0	No	0
3483	80	Male	104	3757	Yes	No	0.0	No	1
21131	70	Female	0	4	Yes	No	0.0	No	0
21131	70	Female	4	22	Yes	Yes	15.0	Yes	0
21131	70	Female	22	64	Yes	Yes	5.0	Yes	0
21131	70	Female	64	73	Yes	Yes	15.0	Yes	0
21131	70	Female	73	120	Yes	Yes	2.5	Yes	0
21131	70	Female	120	154	Yes	Yes	15.0	Yes	0
21131	70	Female	154	165	Yes	Yes	15.0	No	1

MULTIPLE IMPUATION OF GLUCOCORTICOID DOSE AND COVARIATES

Imputation of exposure and covariate missing data took place in 2 consecutive steps.

Step 1: Multiple imputation of glucocorticoid dose

During the study follow-up, the median number of prescriptions recorded per patient was 25 (IQR 11-56). The median time between prescriptions was 31 days (IQR 22-56). The quantity prescribed (i.e. total number of tablets) was available for all patients, so we were able to calculate the duration of each prescription. In addition, we allowed a grace period of 90 days

between refills to account for ≥2 prescriptions as being the same episode of GC dose exposure. The median number of prescriptions with missing dose per patient (e.g. which generally corresponds to tapering periods) was 6.3% (IQR 0.1-41.7%).

Missing daily glucocorticoid dose appeared to be missing at random after adjusting for major confounders (e.g. age, underlying disease, disease duration). It was therefore replaced through generation of 5 datasets using multiple imputation with chained equations (MICE package in R 3.3.1).

The imputation model specifications and variable missingness are shown in the table below.

Variable	**Variable Type**	**Missing (%)**	**Imputation method**
Patient indicator	Continuous, non-normal	0	Predictor/Auxiliary variable
Family practice indicator	Continuous, non-normal	0	Predictor/Auxiliary variable
Age	Continuous, non-normal	0	Predictor/Auxiliary variable
Sex	Binary	0	Predictor/Auxiliary variable
Ethnicity	Category	8.1	Polytomous logistic regression
Index of multiple deprivation	Continuous, non-normal	0	Predictor/Auxiliary variable
Underlying inflammatory disease (PMR, GCA, or both)	Category	0	Predictor/Auxiliary variable
Duration of underlying inflammatory disease	Category	0	Predictor/Auxiliary variable
Daily oral prednisolone-equivalent glucocorticoid dose	Continuous, non-normal	45.7*	Predictive mean matching
Type of oral glucocorticoid**	Category	0	Predictor/Auxiliary variable
Time between the follow-up start and the date of the first glucocorticoid period	Continuous, non-normal	0	Predictor/Auxiliary variable
CVD	Binary	0	Predictor/Auxiliary variable
Diabetes	Binary	0	Predictor/Auxiliary variable
Prescribed non-oral GCs	Binary	0	Predictor/Auxiliary variable

Abbreviations: CVD, cardiovascular disease; DMARDs, disease-modifying antirheumatic drugs; GCA, giant cell arteritis; NSAIDs, non-steroidal anti-inflammatory drugs; PMR, polymyalgia rheumatica

* Please note that this is the total percentage of missing dose in relation to the total number of prescriptions, not in relation to the total number of patients or to the total number of glucocorticoid exposure episodes/periods. As mentioned previously, the median number of prescriptions with missing dose per patient was 6.3%.

** Oral glucocorticoid drugs were: prednisolone, prednisone, budenoside, beclomethasone, betamethasone, deflazacort, dexamethasone, hydrocortisone, triamcinolone, cortisone and methylprednisolone

Five multiply imputed datasets were generated. The density of the imputed data for each imputed dataset was compared against the density of the observed data. The Kolmogorov-Smirnov test was also used to compare the difference between the observations and imputed values

Step2: Multiple imputation of covariates for adjustment

We performed the second imputation step for each of the 5 dose-imputed datasets separately. Risk factor data appeared to be missing at random after adjusting for major confounders (e.g. age, sex, and underlying disease conditions). Hence, multiple imputation via chained equations was implemented using the MICE package in R 3.3.1, to replace missing values in risk factor variables. The imputation model specifications and variable missingness are shown in the table below.

Variable	Variable Type	Missing (%)	Imputation method
Age	Continuous, non-normal	0	Predictor/Auxiliary variable
Sex	Binary	0	Predictor/Auxiliary variable
Ethnicity	Category	0 (Imputed)	Polytomous logistic regression
Underlying disease conditions	Category	0	Predictor/Auxiliary variable
BMI	Continuous, non-normal	57.7	Predictive mean matching
Index of multiple deprivation	Continuous, non-normal	0	Predictor/Auxiliary variable
Number of hospital admissions	Continuous, non-normal	8.9	Predictive mean matching
Daily dose	Continuous, non-normal	0 (Imputed)	Predictor/Auxiliary variable
Smoking status	Binary	16.2	Default imputation (missing set to no)
CVD	Binary	0	Predictor/Auxiliary variable
Diabetes	Binary	0	Predictor/Auxiliary variable
Cancer	Binary	0	Predictor/Auxiliary variable
Asthma	Binary	0	Predictor/Auxiliary variable
COPD	Binary	0	Predictor/Auxiliary variable
Renal disease	Binary	0	Predictor/Auxiliary variable
HIV	Binary	0	Predictor/Auxiliary variable
Influenza vaccination status	Binary	0	Predictor/Auxiliary variable
Varicella zoster vaccination	Binary	0	Predictor/Auxiliary variable
Pneumococcus vaccination	Binary	0	Predictor/Auxiliary variable
Prescribed non-oral GCs	Binary	0	Predictor/Auxiliary variable
H2 antagonists	Binary	0	Predictor/Auxiliary variable

Variable	Variable Type	Missing (%)	Imputation method
Proton pump inhibitors	Binary	0	Predictor/Auxiliary variable
DMARD	Binary	0	Predictor/ Auxiliary variable
NSAIDS	Binary	0	Predictor/ Auxiliary variable

Five multiply imputed datasets were generated for each of the 5 dose-imputed datasets. The density of the imputed data for each imputed dataset was compared against the density of the observed data. The Kolmogorov-Smirnov test was also used to compare the difference between the observations and imputed values.

Appendix to: Wu J, Keeley A, Mallen C, et al. Incidence of infections associated with oral glucocorticoid dose in people diagnosed with polymyalgia rheumatica or giant cell arteritis: a cohort study in England. *CMAJ* 2019. doi: 10.1503/cmaj.190178.

REFERENCES

1. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International journal of epidemiology.* 2015;44:827-36.
2. Mathur R, Bhaskaran K, Chaturvedi N, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J Public Health (Oxf).* 2014;36:684-92.
3. Meier CR, Napalkov PN, Wegmuller Y, Jefferson T, Jick H. Population-based study on incidence, risk factors, clinical complications and drug utilisation associated with influenza in the United Kingdom. *Eur J Clin Microbiol Infect Dis.* 2000;19:834-42.
4. Ronquist G, Rodriguez LA, Ruigomez A, et al. Association between captopril, other antihypertensive drugs and risk of prostate cancer. *Prostate.* 2004;58:50-6.
5. Ryan R, Majeed A. Prevalence of treated hypertension in general practice in England and Wales, 1994 and 1998. *Health Stat Q.* 2002;16:14-8.
6. van Staa TP, Dennison EM, Leufkens HG, Cooper C. Epidemiology of fractures in England and Wales. *Bone.* 2001;29:517-22.

Appendix to: Wu J, Keeley A, Mallen C, et al. Incidence of infections associated with oral glucocorticoid dose in people diagnosed with polymyalgia rheumatica or giant cell arteritis: a cohort study in England. *CMAJ* 2019. doi: 10.1503/cmaj.190178.