

Supplementary file

Data sources

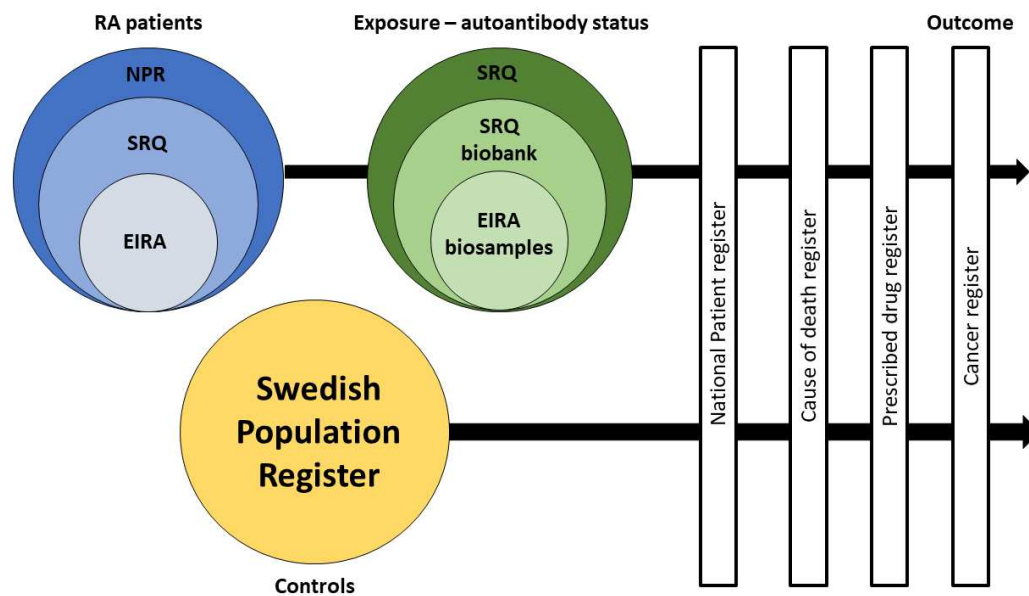
Sweden has national and virtually complete registers on demographics and health data that can be linked together by the unique personal identity number issued to all Swedish residents. This study was based on linkages between the Swedish Rheumatology Quality Register (SRQ) and its adjunct biobank (SRQ biobank), the Swedish Epidemiological Investigation of RA (EIRA) case-control study, and the National Patient Register (NPR), the Swedish Cancer Register, the Prescribed Drug Register, the Swedish Cause of Death Register and the Swedish Population Register (supplementary figure 1). These data sources have been described elsewhere [Askling J, Fored CM, Geborek P, *et al.* Swedish registers to examine drug safety and clinical issues in RA. *Ann. Rheum. Dis.* 2006].

In brief, the SRQ collects clinical data on incident and prevalent patients with RA since 1995. Since 2012, the SRQ biobank collects blood samples from patients included in SRQ (incident patients, and patients who have started a bDMARD). EIRA is a population-based case-control study of incident RA and matched population controls from defined geographical areas of central Sweden, initiated in 1996 and still ongoing [Stolt P, Bengtsson C, Nordmark B, *et al.* Quantification of the influence of cigarette smoking on rheumatoid arthritis: Results from a population based case-control study, using incident cases. *Ann Rheum Dis* Published Online First: 2003].

The Swedish Cancer Register covers >95% of incident primary cancers and contains data on cancer date, cancer-morphology and type using the ICD (International Classification of Diseases) 7 through 10 and ICD-O classifications [Barlow L, Westergren K, Holmberg L, *et al.* The completeness of the Swedish Cancer Register - A sample survey for year 1998. *Acta Oncol (Madr)* Published Online First: 2009]. The National Patient Register covers hospital discharges since 1987 [Ludvigsson JF, Andersson E, Ekbom A, *et al.* External review and validation of the Swedish national inpatient register. *BMC Public Health* Published Online First: 2011.] and outpatient

visits in specialised care (including rheumatology) since 2001. The Swedish Cause of Death Register provides information on dates and causes of death for all deceased residents nationwide since 1961. The Swedish Population Register includes data on residency and dates of immigration and emigration for residents in Sweden since 1961 and onwards, with almost complete coverage.

Supplementary Figure 1. Registers and other data sources used in this study

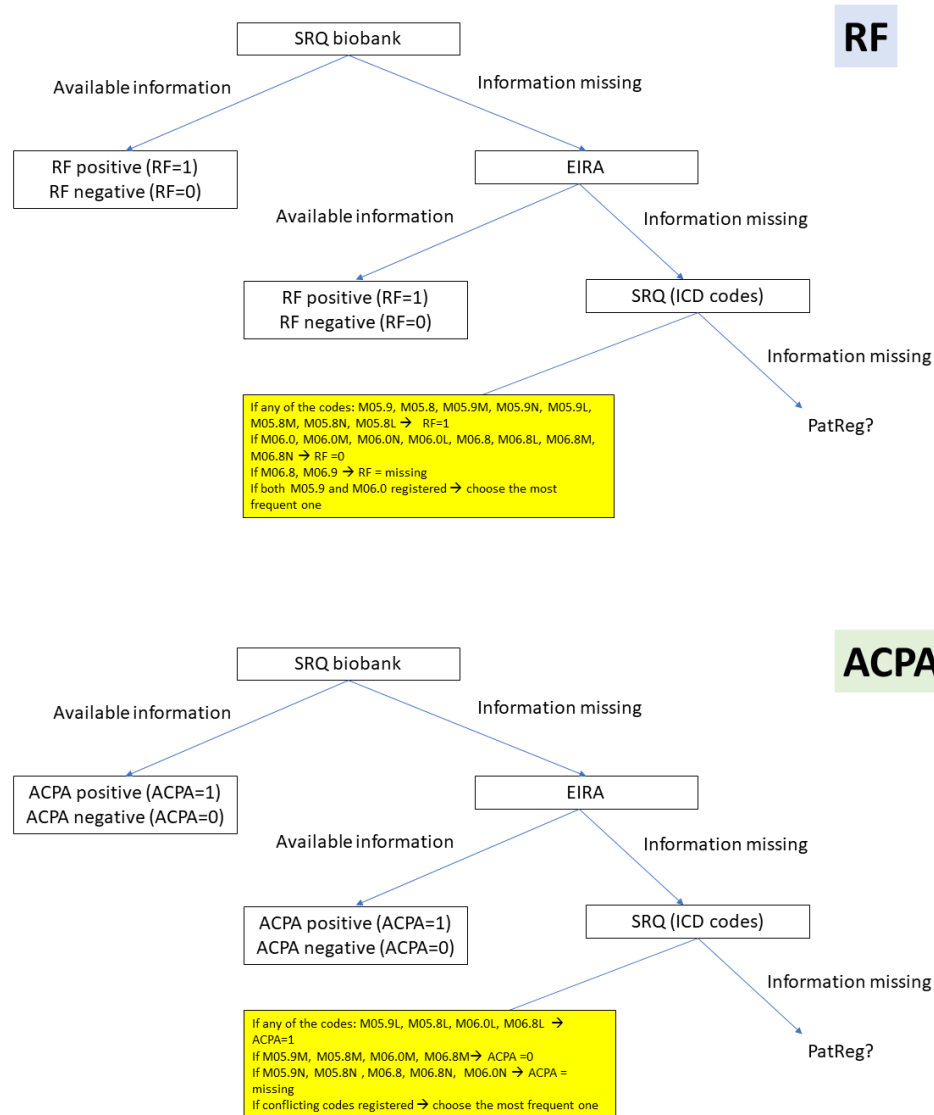


RA = Rheumatoid Arthritis; NPR = National Patient Register; SRQ = Swedish Rheumatology Quality Register; EIRA = Swedish epidemiological investigation of RA

Study population

We defined our study cohort of incident RA from the following three overlapping data sources: a) All patients with a rheumatologist-assigned diagnosis of RA ICD (International Classification of Disease) code M05 and M06 including subheadings in SRQ (n=18,902); b) All patients with RA included in the EIRA study 1995 or later (N=2,060); c) All patients with incident RA in the Swedish Patient Register, here defined as two or more first-ever visits or hospitalisations listing RA (ICD10 M05 or M06 and all their subheadings), at least one of which from a department of internal medicine or rheumatology, 2006 or later, thus allowing a five-year washout of prevalent cases since the inception of the Patient Register in 2001 (n=23,139). Individuals receiving (according to SRQ or to information in the Prescribed Drug Register) prescriptions for methotrexate, sulfasalazine, tumour necrosis factor inhibitors, abatacept, or tocilizumab more than 6 months before their first visit listing RA were excluded. For individuals included in more than one of the three cohorts a) through c), we used the earliest registered date to define cohort entry. In total, 44,101 patients with RA were included.

For each unique patient with RA, we identified five matched general population comparator subjects, alive the year of entry of their index individual (entry here being defined as the year of fulfilment of the criteria for entry into each of the above cohorts). Comparator subjects were matched for age, sex, calendar period and area of residence, and received the same date of entry as their matched index-case with RA. For patients with RA identified in the EIRA study we also included the original matched (1:1-1:2) population-controls included in that study. In the sub-analyses restricted to EIRA patients, we used the original matched population controls and the date of inclusion in EIRA as start of follow-up.

Supplementary figure 2A and 2B: Algorithm for hierarchical definition of RF and ACPA positivity

RF = rheumatoid factor, ACPA = anti-citrullinated peptide antibodies, SRQ = Swedish Rheumatology Quality Register, EIRA = Swedish epidemiological investigation of RA, ICD = International Classification of Diseases

Supplementary Table 2: Information about comorbidity, up to 5 years before RA diagnosis

Outcome	ICD 10	ICD 9	ICD 8	Data source
COPD	J40-J44	490 , 491.0, 491.1, 491.2, 491.8, 492, 496	490 , 491.01, 491.02, 491.04, 492	Patient Register, visits (main dx), hospitalisation (main dx, but if main dx is M05* or M06* then bidiagnos 1 is searched for the ICD 10 codes
All respiratory tract diseases	J00-J99	460-519	460-519	Patient Register, visits (main dx), hospitalisation (main dx, but if main dx is M05* or M06* then bidiagnos 1 is searched for the ICD 10 codes
All respiratory tract diseases except infections	J30-J99	470-478, 490-519	485, 490-519	Patient Register, visits (main dx), hospitalisation (main dx, but if main dx is M05* or M06* then bidiagnos 1 is searched for the ICD 10 codes
Acute upper respiratory tract diseases	J00-J06, J36	460-465, 475	460-464, 501	Patient Register, visits (main dx), hospitalisation (main dx, but if main dx is M05* or M06* then bidiagnos 1 is searched for the ICD 10 codes
Chronic upper respiratory tract diseases	J30-J35, J37-J39	470-474, 476-478	500, 502-508	Patient Register, visits (main dx), hospitalisation (main dx, but if main dx is M05* or M06* then bidiagnos 1 is searched for the ICD 10 codes
Acute lower respiratory tract diseases	J09-J18, J20-J22, J69	466, 480-487, 507	466 , 470-474, 480-486	Patient Register, visits (main dx), hospitalisation (main dx, but if main dx is M05* or M06* then bidiagnos 1 is searched for the ICD 10 codes
Chronic lower respiratory tract diseases	J40-J47, J60-J64, J66-67, J82, J84	490-495, 500- 505, 515-516, 518.3	490-493, 515- 518, 519.20, 519.21, 519.22	Patient Register, visits (main dx), hospitalisation (main dx, but if main dx is M05* or M06* then bidiagnos 1 is searched for the ICD 10 codes
Asthma	J45-J46	493	493	
Hospitalized Infections	A00-B99, D73.3, E06.0, E32.1, G00- G02, G04.2, G05- G07, H00.0, H44.0, H60.0-H60.3, H66- H67, H70, I30.1, I40.0, J00-J22, J32, J34.0, J36, J38.3, J39.0-J39.1, J44.0, J85, J86, K04.4, K04.6, K04.7,			Main diagnosis in inpatient component of Patient Register. If main diagnosis was RA (ICD10 codes M05, M06.0, M06.2, M06.3, M06.8, M06.9, M12.3), contributory diagnoses of hospitalized infections were also allowed. Note that infection data from non-primary (specialist)

	K10.2, K11.3, K12.2, K14.0, K57.0, K57.2, K57.4, K57.8, K61, K63.0, K65.0, K65.1, K65.2, K65.9, L00-L08, L30.3, M00-M01, M46.2-M46.5, M60.0, M65.0, M71.0, M71.1, M72.6, M86, N10, N11, N12, N13.6, N15.1, N15.9, N30.0 N30.8, N34.0, N41.2, N43.1, N45.2, N45.3, N45.4, N48.2, N61, N70, N73, N75.1	outpatient care was not included since it would be highly influenced by surveillance effects. In particular, patients seen for a chronic disease would in passing be more likely to also be recorded with these milder conditions. Regardless, most non-serious infections would not be treated in specialist care.
Malignancy excluding non-melanoma and basal cell invasive/non-invasive skin cancers	All non-benign tumors, except C44 and D04 (ICD7=191), and basal cell cancers	The Cancer Register
Non-melanoma invasive/non-invasive skin cancers including basal cell cancer.	C44 and D04 (ICD7=191)	The Cancer Register
Heart failure	I50	Patient Register, visits (main dx), hospitalisation (main dx, but if main dx is M05* or M06* then bidiagnos 1 is searched for the ICD 10 codes
Ischemic heart disease	I20-I25	Patient Register, visits (main dx), hospitalisation (main dx, but if main dx is M05* or M06* then bidiagnos 1 is searched for the ICD 10 codes
Renal failure	N18-19	Patient Register, visits (main dx), hospitalisation (main dx, but if main dx is M05* or M06* then bidiagnos 1 is searched for the ICD 10 codes
All-cause hospitalization	-	The inpatient component of the Patient Register
All-cause mortality	-	The Cause of Death Register