Scandinavian Journal of Rheumatology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/irhe20

The validity of rheumatoid arthritis diagnoses in Finnish biobanks

J Paltta, H-K Heikkilä, L Pirilä, KK Eklund, J Huhtakangas, P Isomäki, O Kaipiainen-Seppänen, K Kristiansson, AS Havulinna, T Sokka-Isler, A Palomäki & for the FinnGen investigators

To cite this article: J Paltta, H-K Heikkilä, L Pirilä, KK Eklund, J Huhtakangas, P Isomäki, O Kaipiainen-Seppänen, K Kristiansson, AS Havulinna, T Sokka-Isler, A Palomäki & for the FinnGen investigators (2021): The validity of rheumatoid arthritis diagnoses in Finnish biobanks, Scandinavian Journal of Rheumatology, DOI: 10.1080/03009742.2021.1967047

To link to this article: https://doi.org/10.1080/03009742.2021.1967047

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

View supplementary material

Published online: 13 Oct 2021.

Submit your article to this journal

Article views: 229

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=irhe20

The validity of rheumatoid arthritis diagnoses in Finnish biobanks

J Paltta[1], H-K Heikkilä[2], L Pirilä[1], KK Eklund[3], J Huhtakangas[4], P Isomäki[2,5], O Kaipiainen-Seppänen[6], K Kristiansson[7], AS Havulinna[7,8], T Sokka-Isler[9], A Palomäki[1,8] for the FinnGen investigators[10]

[1]Centre for Rheumatology and Clinical Immunology, Division of Medicine, Turku University Hospital and University of Turku, Turku, Finland
[2]Centre for Rheumatic Diseases, Tampere University Hospital, Tampere, Finland
[3]Department of Rheumatology, Helsinki University Hospital, University of Helsinki and Orton Orthopaedic Hospital, Helsinki, Finland
[4]Division of Rheumatology, Department of Internal Medicine, Oulu University Hospital and Medical Research Center Oulu, Oulu, Finland
[5]Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
[6]Department of Medicine, Kuopio University Hospital, Kuopio, Finland
[7]Department of Public Health Solutions, Finnish Institute for Health and Welfare (THL), Helsinki, Finland
[8]Institute for Molecular Medicine Finland, FIMM, HiLIFE, University of Helsinki, Helsinki, Finland
[9]Department of Medicine, Jyväskylä Central Hospital, Jyväskylä, Finland
[10]FinnGen members are listed in the Supplementary material

Objective: The aim of this study was to determine the validity of rheumatoid arthritis (RA) diagnoses in patients participating in Finnish biobanks.

Method: We reviewed the electronic medical records of 500 Finnish biobank participants: 125 patients with at least one visit with a diagnosis of seropositive RA, 125 patients with at least one visit with a diagnosis of seronegative RA, and 250 age- and gender-matched controls. The patients were chosen from five different biobank hospitals in Finland. A rheumatologist reviewed the medical records to assess whether each patients' diagnosis was correct. The diagnosis was compared with the diagnostic codes in the Finnish Care Register for Health Care (CRHC) and special reimbursement data of the Social Insurance Institution of Finland.

Results: The positive predictive value (PPV) of CRHC diagnosis of RA (for seropositive and seronegative RA combined) was 0.82. For patients with a special reimbursement for anti-rheumatic medications for RA, the PPV was 0.89. The PPV was higher in patients with more than one visit. For one, two, five, and 10 visits, the PPV was 0.82, 0.85, 0.89, and 0.90, respectively, and for patients who also had the special reimbursement, the PPV was 0.89, 0.91, 0.93, and 0.94 for one, two, five, and 10 visits, respectively. In patients positive for anti-citrullinated protein antibodies, the PPV was 0.98.

Conclusion: These results demonstrate that the validity of RA diagnoses in Finnish biobanks was good and can be further improved by including data on special reimbursement for medication, number of visits, and serological data.

Biobanks are collections of biological samples and related health data (1). In biobank-based studies, data obtained from collected biological samples are combined with related data from electronic health records (2, 3). Biobank-based research often relies on diagnostic information recorded in healthcare registers (3). The information in these registers is collected in daily clinical practice and not primarily for research purposes (4). Information about the accuracy of the healthcare registers is essential when designing studies based on these registers and evaluating the results of register-based studies. Biobank studies also rely on diagnostic information from medical records and registers for selecting patients for research cohorts.

In Finland, the Care Register for Health Care (CRHC) contains information about patients visiting specialized outpatient care or discharged from inpatient care in all hospitals in Finland. The diagnoses of these contacts have been recorded using the 10th revision of the International Classification of Diseases (ICD-10) since 1996.

The validity of CRHC diagnoses has been previously studied in certain disease groups (5, 6), with a generally high accuracy in the diagnoses. Clearly fewer studies have been conducted on the accuracy of diagnostic information in Finnish biobank patients (7) and, to our

Johanna Paltta, Centre for Rheumatology and Clinical Immunology, Division of Medicine, Turku University Hospital and University of Turku, Kiinamyllynkatu 4-8, PL 52, Turku, 20521 Finland.
E-mail: johanna.paltta@tyks.fi

Accepted 9 August 2021

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

DOI: 10.1080/03009742.2021.1967047

knowledge, there are no validation studies of rheumatoid arthritis (RA) diagnoses in Finnish biobank patients or in the CRHC registry.

The aim of the study was to analyse the accuracy of RA diagnoses in Finnish biobank patients. We also studied how this diagnostic information corresponds to the information recorded in the Finnish national healthcare registers, focusing on the CRHC and special reimbursement registry for medication of the Social Insurance Institution of Finland. We also explored whether the accuracy of the diagnostic information would improve by combining data from different healthcare registers.

Method

Study population

The patients included in this study were selected from the records of five hospital biobanks in Finland (the Auria Biobank in Turku, Finland; the Finnish Clinical Biobank Tampere in Tampere, Finland; the Biobank Borealis in Oulu, Finland; the Biobank of Eastern Finland in Kuopio, Finland; and the Central Finland Biobank in Jyväskylä, Finland). The study sample included 125 patients with a diagnosis of seropositive RA (ICD-10 codes M05.8 and M05.9, 25 patients from each biobank), and 125 patients with a diagnosis of seronegative RA (ICD-10 code M06.0, 25 patients from each biobank). The control group consisted of 250 age- and gender-matched controls with 50 patients from each biobank, who had no diagnosis of RA registered in the patient records of the participating hospitals.

The participating biobanks made the initial random selection of patients using the diagnostic codes registered in the participating hospitals' electronic medical records for the years 2007–2018. The patients were selected for the study if they had at least one visit with the inclusion diagnosis to the hospital during these years. The initial diagnosis of RA could have been made before these years or in another hospital, but at least one visit with RA diagnosis during the study years to the biobank hospital was required. The year of the initial diagnosis was collected from the patient records by the reviewer. Five patients in the control group were later found to have visits with a diagnosis of RA registered in the CRHC from another hospital not participating in this study. These five patients were excluded from the control group. Three patients with a diagnosis of seropositive RA and two patients with a diagnosis of seronegative RA recorded in local hospital patient records were later found to have no visits with these diagnoses in the CRHC and thus were analysed as a part of the control group. For four patients in the seropositive group and for eight patients in the seronegative group, there were insufficient electronic patient record data available in the participating biobank hospital for analysis, which resulted in the reviewer not being able to form an expert opinion about the correctness of the diagnosis. These 12 patients were excluded from the final analysis.

Patients treated in the biobank hospitals have the option to include their data in the biobank during any hospital visit. According to Finnish biobank legislation, written consent is obtained from each patient before his or her data are included. Inclusion could take place at any time during the treatment of RA or another medical condition.

Register data

Information about patient visits with a diagnosis of seropositive or seronegative RA was obtained from the CRHC, which, prior to 1994, was called the Hospital Discharge Register, and this is maintained by Finnish Institute for Health and Welfare. The CRHC contains connected data, nationwide, on all hospital inpatient discharges through a personal identification code since 1969 and also with outpatient visits to hospitals since 1998 (5). The register data were obtained after the initial screening of the patients from regional biobanks. This resulted in some patients being assigned to the control group but later excluded from the final analysis owing to a diagnosis of RA in another hospital, and some patients being assigned to the RA group but later analysed as controls, if no diagnosis of RA was found in the CRHC.

The Finnish national health insurance system entitles all patients with certain chronic and severe diseases, such as RA, to special reimbursement for the costs of medications. Information about these reimbursement entitlements and purchases of the medications is recorded in a register maintained by the Social Insurance Institution of Finland. From this medical reimbursement register, we searched for whether the patient had been granted entitlement to reimbursement of the cost of disease-modifying anti-rheumatic drugs [DMARDs; reimbursement with the code 202 for connective tissue diseases (CTDs), RA, and comparable diseases] or entitlement to reimbursement of the cost of biological disease-modifying anti-rheumatic drugs (bDMARDs; reimbursement with the codes 281 or 313 for RA, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis, and comparable diseases). We also analysed whether these reimbursement entitlements had been granted for seropositive RA (ICD-10 M05) or seronegative RA (ICD-10 M06), specifically.

Clinical data

The clinical data were collected from the electronic medical records of the participating biobank hospitals.

The chart review was carried out by a rheumatologist (JP, JH, OK, TS) or an experienced resident in

rheumatology (HH). The reviewer evaluated the correctness of the RA diagnosis according to a thorough chart review and a complete clinical follow-up. This reviewer confirmed a true positive diagnosis of RA based on whether the patient had been diagnosed with RA by an internist, a rheumatologist, or a resident working at a rheumatology clinic, or whether a rheumatologist had confirmed the diagnosis made elsewhere; whether the patient was treated with DMARDs for RA; and whether the complete clinical follow-up was compatible with RA. Fulfilment of American College of Rheumatology (ACR) 1987 or ACR/European League Against Rheumatism (EULAR) 2010 classification criteria for RA was not required (8, 9).

Collected data included symptoms and clinical findings, laboratory and imaging results, and information on the medication used by the patients. For some patients, the exact numeric value of their rheumatoid factor (RF) or anti-citrullinated protein antibody (ACPA) laboratory result was not available, either because the patient had been diagnosed elsewhere or because the diagnosis was from the time before electronic medical records, but in all of those cases, the patients' serostatus could be confirmed from the written medical records. Whether the patient fulfilled the ACR 1987 and ACR/EULAR 2010 classification criteria for RA was determined. It was also recorded whether the diagnosis had been made at a rheumatology clinic or in another healthcare unit, and also whether the diagnosis was erroneously recorded, and some other disease would better explain the patients' symptoms and findings during the follow-up. For controls, it was evaluated whether the medical records showed evidence of the patient having RA without a formal ICD-10 code recorded in the CRHC and being categorized as 'false negative'.

Study data were collected and managed using REDCap electronic data capture tools, hosted at the University of Turku (10, 11).

Statistics

All statistical analyses were performed using R version 3.6.2 with The R base, tidyr, epiR, stats, dplyr, descr, and vcd packages. Continuous variables are expressed as medians with interquartile ranges, and categorical variables are described as counts with percentages.

The positive predictive values and negative predictive values (PPVs and NPVs), positive likelihood ratios and negative likelihood ratios (PLRs and NLRs), and diagnostic accuracy for CRHC diagnosis of RA concurring with a clinical diagnosis of RA were calculated. Information about the patients' entitlement to reimbursement of the cost of medicines and for a positive test for ACPA was also taken into consideration when available. The Cohen's kappa was calculated to measure the agreement between the CRHC diagnosis of RA and the clinical diagnosis of RA. Exact 95% confidence intervals (CIs) were calculated for all of the predictive statistics.

In addition to analysing patients with seropositive RA as a group and patients with seronegative RA as a group, we analysed all the patients with either seropositive or seronegative RA as a group, being 'all RA'. This made our study more comparable to previous studies (12–14), which have not separated seropositive and seronegative subgroups of RA.

Ethical considerations and study permissions

This was a non-interventional retrospective study without any direct patient contact, and according to Finnish legislation, no patient consent or ethical committee approval was needed. The Ethical Committee of Hospital District of Southwest Finland was still consulted, and the committee found no ethical problems in this study (Dnro 62/1804/2019). Permissions for the study were obtained from the Finnish Institute for Health and Welfare (permission no THL/1233/5.05.00/2019), the Social Insurance Institution of Finland (no 77/522/2019), the hospital district, and the biobank of every hospital contributing to the study. The legal basis for processing personal data is public interest and scientific research [EU General Data Protection Regulation 2016/679 (GDPR), Article 6(1)(e) and Article 9(2)(j); Data Protection Act, Sections 4 and 6].

Results

The inclusion and exclusion criteria of the patients are shown in Figure 1. In the final analysis, there were 118 patients with seropositive RA and 115 patients with seronegative RA. Characteristics of the study population are presented in Table 1. A majority of the patients were diagnosed (90%) and treated (95%) at a rheumatological department. The age at diagnosis was younger in patients with seropositive RA, at 46 years, than in patients with seronegative RA, at 54 years. All patients with seropositive RA were treated with DMARDs compared to 95% of the patients with seronegative RA. Moreover, 95% of all RA patients were entitled to reimbursement for DMARDs, and for 66% of the patients, this entitlement had been granted specifically for RA. The analysis found that 63% of the patients with seropositive RA and 30% with seronegative RA had developed radiographic changes suggestive of RA by the end of the follow-up period.

Seropositive and seronegative RA patients

If a patient had at least one visit with a diagnosis of either seropositive or seronegative RA in the CRHC, the

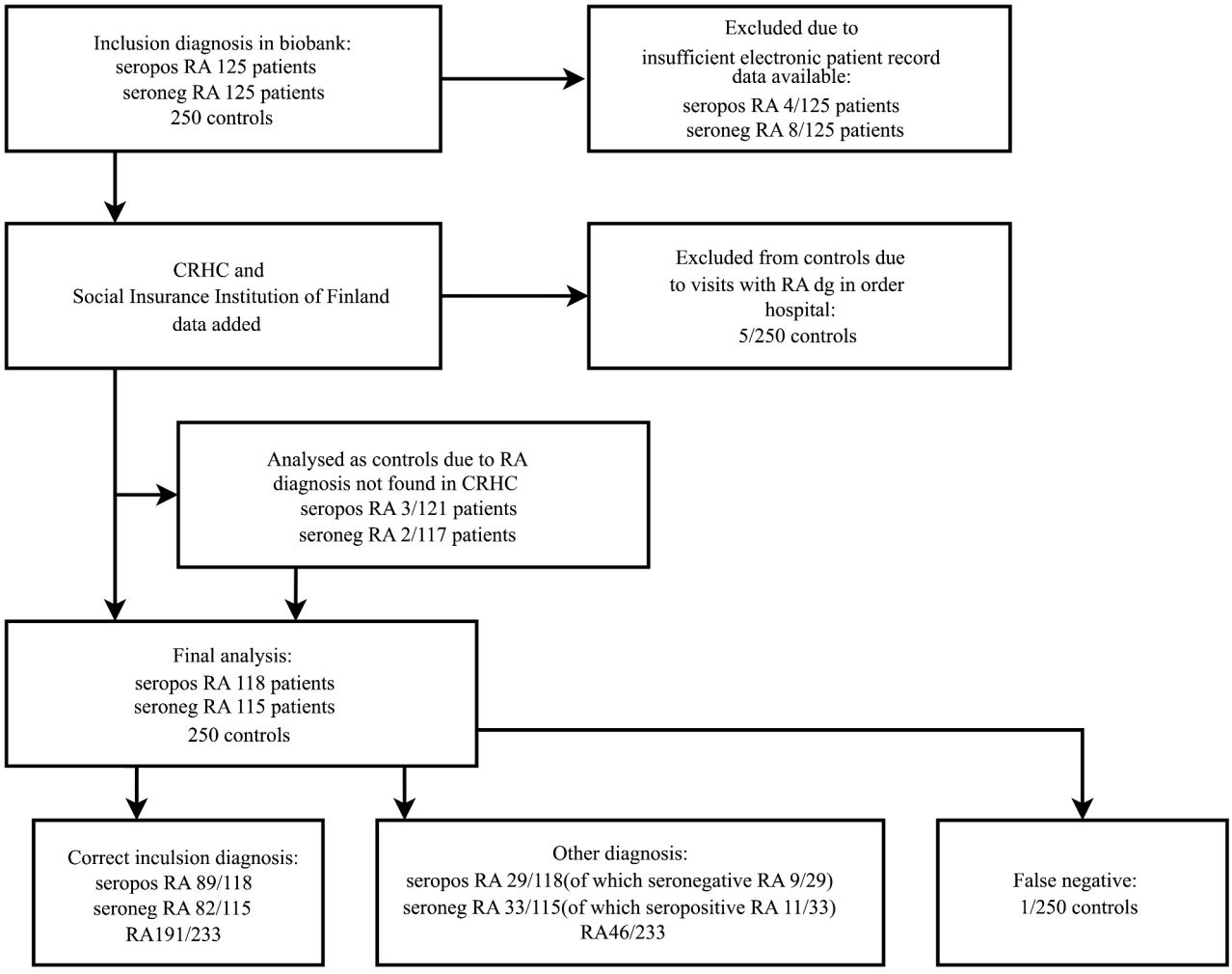

Figure 1. Study flowchart. RA, rheumatoid arthritis; CRHC, Finnish Care Register for Health Care; seropos, seropositive; seroneg, seronegative; dg, diagnosis.

PPV for a diagnosis of RA was 0.82 (191/233). The PPV rose with the number of visits and was 0.85 (189/222) for two visits, 0.89 (175/197) for five visits, and 0.9 (137/152) for 10 visits (Figure 2, Table 2).

For patients with entitlement to reimbursement for DMARDs with a diagnosis of either seropositive or seronegative RA specifically, the PPV for a single visit was 0.89 (136/152) and this value grew higher with two visits (PPV = 0.91, 136/150), five visits (PPV = 0.93, 126/136), and 10 visits (PPV = 0.94, 98/104) (Figure 3, Table 2). With entitlement to reimbursement for the cost of DMARDs with a less specific code of 202 (e.g. CTDs, RA, and comparable diseases), the PPV was 0.84 (183/219) for a single visit, 0.87 (182/210) for two visits, 0.9 (168/187) for five visits, and 0.9 (133/147) for 10 visits (Supplementary figure S1).

In some biobank studies, laboratory values have been included in the criteria for patient selection. In the present study, if the ACPA status was included, the PPV for diagnosis of RA for ACPA-positive patients was 0.98 (62/63) (Table 2).

We found that 92% (160/173) of the patients who were categorized as having RA fulfilled either the ACR 1987 or ACR/EULAR 2010 classification criteria for RA.

Seropositive patients

If a patient had at least one visit with a diagnosis of seropositive RA in the CRHC, the PPV was 0.75 (89/118). The PPV increased with the number of visits, being 0.8 (88/110) with two visits, 0.85 (82/96) with five visits, and 0.91 (70/77) with 10 visits (Figure 2, Table 2).

For the patients having reimbursement for DMARDs with a diagnosis of seropositive RA specifically, the PPV was 0.93 (57/61) if they had a single visit, 0.93 (57/61) if they had two visits, 0.96 (53/55) if they had five visits, and 0.96 (44/46) with 10 visits (Figure 3, Table 2).

Table 1. Demographic and clinical characteristics of the study sample.

	N with data	All RA	Seropositive RA	Seronegative RA
Number of patients	233	233	118	115
Female (%)	233	160 (69%)	75 (64%)	85 (74%)
Year of diagnosis [IQR]	222	2005 [1996–2013]	2001 [1988–2012]	2008 [2001–2013]
Age at diagnosis in years [IQR]	224	50.0 [40.0–59.0]	46.0 [36.0–56.0]	54.0 [45.0–61.1]
Follow-up in years [IQR]	220	11.0 [4.3–22.0]	16.0 [4.5–30.0]	9.0 [4.0–16.0]
Diagnosed in rheumatology (%)	188	169 (90%)	79 (91%)	90 (89%)
Treated in rheumatology (%)	228	209 (95%)	103 (94%)	106 (95%)
Nr. of visits seropositive RA [IQR]	233	5.0 [0.0–25.0]	20.0 [6.0–31.5]	0 [0.0–3.0]
Nr. of visits seronegative RA [IQR]	233	2.0 [0.0–13.0]	0.0 [0.0–0.0]	11.0 [4.5–23.0]
Treated with DMARDs (%)	221	215 (97%)	110 (100%)	105 (95%)
Reimbursement for DMARDs (inclusion diagnosis specific) (%)	230	152 (66%)	61 (52%)	75 (66%)
Reimbursement for DMARDs (%)	230	219 (95%)	111 (95%)	108 (96%)
EULAR classification criteria positive at diagnosis (%)	136	79 (58%)	47 (72%)	32 (45%)
ACR classification criteria positive at diagnosis (%)	128	78 (61%)	41 (67%)	37 (55%)
ACR or EULAR classification criteria positive at diagnosis (%)	128	95 (74%)	49 (80%)	46 (69%)
EULAR classification criteria positive ever (%)	197	145 (74%)	89 (86%)	56 (60%)
ACR classification criteria positive ever (%)	176	131 (74%)	73 (77%)	58 (72%)
ACR or EULAR classification criteria positive ever (%)	195	164 (84%)	90 (89%)	74 (79%)
Highest RF ever [IQR]	189	13.0 [6.0–73.0]	75.5 [16.5–205.0]	9.0 [0.0–13.5]
Highest ACPA ever [IQR]	166	1.45 [0.0–126.0]	129.0 [7.0–340.0]	0.8 [0.0–1.6]
ACPA positive ever (%)	166	63 (38%)	55 (75%)	8 (9%)
Erosions in radiographs at diagnosis (%)	143	31 (22%)	19 (32%)	12 (14%)
Erosions in radiographs ever (%)	210	98 (47%)	66 (63%)	32 (30%)

ACPA, anti-citrullinated protein antibody; ACR, American College of Rheumatology; Eular, European League Against Rheumatism; DMARDs, disease-modifying anti-rheumatic drugs; RA, rheumatoid arthritis; RF, rheumatoid factor.

For the patients with visits with a diagnosis of seropositive RA for whom the entitlement to reimbursement for DMARDs had been granted with a less specific code of 202 (e.g. CTDs, RA, and comparable diseases), the PPV for a single visit was 0.79 (88/111), 0.84 (87/104) for two visits, 0.89 (81/91) for five visits, and 0.92 (69/75) for 10 visits (Supplementary figure S1).

In ACPA-positive patients [ACPA greater than the upper limit of normal (ULN)], the PPV for a diagnosis of seropositive RA was 0.96 (53/55) (Table 2). If the ACPA was high positive (three or more times the ULN), the PPV was 0.98 (52/53). If the patient had a diagnosis of seropositive RA and RF was greater than the ULN, the PPV for a diagnosis of seropositive RA was 0.92 (58/63), and if the RF was high positive, being three or more times the ULN, the PPV was 0.94 (49/52).

We found that 97% (83/86) of the patients who were categorized as having seropositive RA fulfilled either the ACR 1987 or ACR/EULAR 2010 classification criteria for RA.

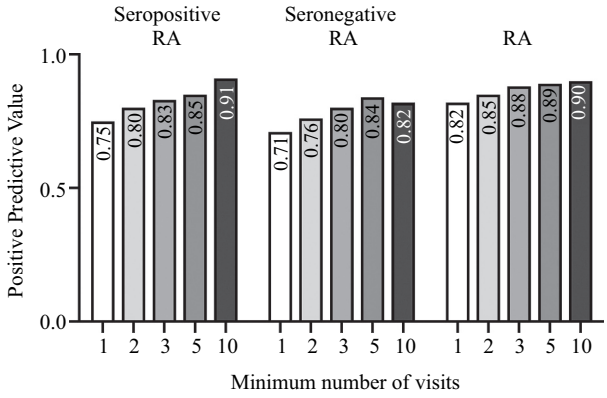

Figure 2. Positive predictive value of rheumatoid arthritis (RA) diagnosis in the Finnish Care Register for Health Care (CRHC) compared to chart review.

Seronegative patients

If a patient had at least one visit with a diagnosis of seronegative RA in the CRHC, the PPV was 0.71 (82/115). The PPV was higher if the patient had two visits (PPV 0.76, 81/106), five visits (PPV 0.84, 72/86), or 10 visits (PPV 0.82, 49/60) (Figure 2, Table 2).

For patients with reimbursement for the cost of DMARDs with a diagnosis of seronegative RA specifically, the PPV was 0.79 (59/75) if the patient had a single visit, 0.82 (59/72) if the patient had two visits, 0.85 (53/62) for five visits, and 0.86 (37/43) for 10 visits (Figure 3, Table 2).

If the patient was entitled to reimbursement for DMARDs with a less specific code 202 (e.g. CTDs, RA, and comparable diseases), the PPV values for patients with a single visit, two visits, five visits, and 10 visits were 0.70 (76/108), 0.76 (76/100), 0.83 (67/81), and 0.81 (47/58), respectively (Supplementary figure S1).

We found that 88% (61/69) of the patients who were categorized as having seronegative RA fulfilled either the ACR 1987 or ACR/EULAR 2010 classification criteria for RA.

Control group

After a thorough examination of the patient records, only one of the 250 controls was found to have information in their patient charts suggesting a diagnosis of RA (NPV 1.0, 249/250). This patient had a diagnosis of seronegative RA recorded in the local hospital, but the diagnosis of RA was not found in the CRHC database.

Incorrect diagnoses

Out of the 233 patients, upon follow-up, 62 patients proved to have an incorrect diagnosis. Of these 62 incorrect diagnoses, in 20 patients, only the seropositivity or seronegativity was considered incorrect. The final diagnosis of the

Table 2. Agreement between register-based diagnoses and chart review.

	All RA	Seropositive RA	Seronegative RA
At least one CRHC visit with RA			
PPV (95% CI)	0.82 (0.76, 0.87)	0.75 (0.67, 0.83)	0.71 (0.62, 0.79)
NPV (95% CI)	1.00 (0.98, 1.00)	1.00 (0.98, 1.00)	1.00 (0.98, 1.00)
PLR (95% CI)	6.89 (5.21, 9.12)	9.48 (6.71, 13.39)	8.44 (6.12, 11.64)
NLR (95% CI)	0.01 (0.00, 0.04)	0.01 (0.00, 0.09)	0.01 (0.00, 0.10)
Accuracy (95% CI)	0.91 (0.88, 0.93)	0.92 (0.89, 0.94)	0.91 (0.87, 0.93)
Kappa (95% CI)	0.82 (0.77, 0.87)	0.80 (0.73, 0.87)	0.77 (0.69, 0.84)
At least one CRHC visit and reimbursement for DMARDs with inclusion diagnosis			
PPV (95% CI)	0.89 (0.83, 0.94)	0.93 (0.84, 0.98)	0.79 (0.68, 0.87)
NPV (95% CI)	1.00 (0.98, 1.00)	1.00 (0.98, 1.00)	1.00 (0.98, 1.00)
PLR (95% CI)	16.44 (10.22, 26.44)	62.16 (23.50, 164.43)	16.29 (10.12, 26.22)
NLR (95% CI)	0.01 (0.00, 0.05)	0.02 (0.00, 0.12)	0.02 (0.00, 0.12)
Accuracy (95% CI)	0.96 (0.93, 0.98)	0.98 (0.96, 0.99)	0.95 (0.92, 0.97)
Kappa (95% CI)	0.91 (0.87, 0.95)	0.95 (0.90, 0.99)	0.84 (0.77, 0.91)
At least one CRHC visit and ACPA positivity			
PPV (95% CI)	0.98 (0.91, 1.00)	0.96 (0.87, 1.00)	
NPV (95% CI)	1.00 (0.98, 1.00)	1.00 (0.98, 1.00)	
PLR (95% CI)	246.03 (34.78, 1740.19)	123.18 (30.96, 490.03)	
NLR (95% CI)	0.02 (0.00, 0.11)	0.02 (0.00, 0.13)	
Accuracy (95% CI)	0.99 (0.98, 1.00)	0.99 (0.97, 1.00)	
Kappa (95% CI)	0.98 (0.95, 1.01)	0.97 (0.93, 1.00)	

RA, rheumatoid arthritis; CRHC, Finnish Care Register for Health Care; PPV, positive predictive value; NPV, negative predictive value; PLR, positive likelihood ratio; NLR, negative likelihood ratio; CI, confidence interval; DMARD, disease-modifying antirheumatic drug; ACPA, anti-citrullinated protein antibody.

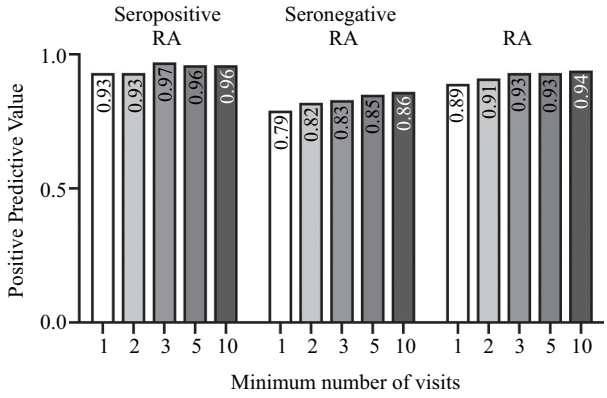

Figure 3. Positive predictive value of rheumatoid arthritis (RA) diagnosis in patients with reimbursement for disease-modifying anti-rheumatic drugs (specifically for RA).

remaining 42 patients included a variety of other rheumatological and unrelated medical conditions, specified in Supplementary table S2.

All (29/29) incorrect diagnoses of seropositive RA and 67% (22/33) of the incorrect diagnoses of seronegative RA were incorrectly input; for example, a diagnosis being a clear deviation from the physician's record and not a misdiagnosis. Of these incorrect recordings, 39% (20/51) were made in the Department of Rheumatology, 24% (12/51) in the Department of Surgery, 8% (4/51) in the Department of Physiotherapy, and the rest 29% (15/51) in various departments. Furthermore, 15% (5/33) of the diagnoses of seronegative RA seemed valid at the time of diagnosis but changed during the follow-up period, and 18% (6/33) of the diagnoses were considered misdiagnoses (Fig-Figure 1).

Discussion

In this study, we validated the RA diagnoses of Finnish biobank patients. The results showed that the PPV for a diagnosis of RA for at least one visit was 0.82. The diagnoses were more accurate if the patient had more than one visit, with, for example, five visits having a PPV of 0.89, or if the patient had entitlement to special reimbursement for anti-rheumatic medications (PPV = 0.89). In some biobank studies, including the present study, laboratory data are available. Inclusion of information on ACPA status clearly increased the accuracy of the diagnosis, and accordingly, the PPV value of a diagnosis of RA for ACPA-positive patients was excellent (0.98).

During recent years, there have been a few studies addressing the accuracy of the Finnish healthcare registers. The meta-analysis of 32 studies by Sund (5) analysed the quality of the Finnish Hospital Discharge Register, which was later replaced by the CRHC. The accuracy of the diagnoses varied between 75% and 95% for common diagnoses. Vuori et al (6) found a PPV of 0.85 (95% CI 0.77–0.91) and a NPV of 0.83 (95% CI 0.75–0.90) for heart failure diagnoses in the Finnish Hospital Discharge Register. Haverinen et al (7) validated psoriasis diagnoses recorded in Finnish biobanks and found a PPV of 88.0% (95% CI 82.7–92.2). To the best of our knowledge, there have been no validation studies in the field of rheumatology in Finland.

In Sweden, Waldenlind et al (12) studied patients who had the diagnosis of RA set by a rheumatology clinic at least on two visits. Approximately 90% of these patients were found to have a definite diagnosis of RA. In Denmark, Ibfelt et al (13) studied the validity of the diagnoses of RA in the DANBIO register, which is a register of inflammatory arthritis diseases, and in the DNPR register, which is the Danish National Patient Register. The inclusion criteria for the patients was either a diagnosis of RA in the DANBIO register or a diagnosis of RA in the DNPR register that was set on at least two visits at a rheumatology clinic. In the DANBIO register, the accuracy of the diagnoses was 96%, and in the DNPR register, it was slightly lower at 79%.

Seronegative RA is a heterogeneous disease entity, and many patients present with a competing diagnosis during follow-up. In a registry study by Paalanen et al, spondyloartrhropathy was diagnosed in 8.8% of patients initially diagnosed with seronegative RA during 15 years of follow-up (15), and in another study with a thorough clinical follow-up, a more specific or competing diagnosis could be proposed even in the majority of patients with seronegative RA during a 10 year follow-up (16). In our study, 15% of incorrect diagnoses of seronegative RA were diagnoses that were changed to a more defined diagnosis during follow-up, and 18% were misdiagnoses. Also, the median hospital follow-up time was shorter in patients with seronegative RA in our study, which may reflect a more self-limiting disease course or more frequent change of diagnosis to a non-rheumatic condition.

Most of the previous studies on RA have validated only the diagnoses that were set at rheumatology clinics. For example, in Minneapolis, USA, the accuracy of the diagnosis of RA in the Veterans Administration database was 55% (14), and in Sweden and in Denmark, the PPV for RA has been reported to be between 0.79 and 0.96 (12, 13). Our study included all RA diagnoses recorded at any clinic, which makes the results more generalizable.

In our study, the validity of the diagnosis was correlated with the number of visits and the patients' entitlement to special reimbursement of the cost of medication. Algorithms previously developed to identify patients with RA from the registers (17–20) have usually included the number and the location, i.e. at a rheumatology clinic or elsewhere, of the diagnoses, anti-rheumatic medications prescribed for the patient, and other rheumatic diagnoses. Our study reinforces the notion that the validity of the diagnoses can be significantly improved by combining data from different registers.

Our study also assessed the impact of serological data, ACPA and RF, on the validity of a diagnosis of RA. In Denmark, Tenstad et al found the PPV of ACPA to be higher than the PPV of RF at a high positivity, being three times the ULN (21). Our results support these findings; the PPVs for a diagnosis of seropositive RA with both positive ACPA and high positive ACPA were higher than the PPVs of a positive RF and a high positive RF.

A limitation of our study was that we had no access to the patient records in other healthcare facilities outside the participating biobank hospitals. This resulted in some patients being assigned to the control group but later excluded from the final analysis, when they were found to have visits with a diagnosis of RA in the CRHC from another hospital in Finland. For the same reason, there were also limited data available for some of the patients in the RA group, resulting in the reviewer not being able to form an expert opinion on the validity of the diagnosis for 12 patients. Nonetheless, the vast majority of the patients had several visits and sufficient data available in the hospital affiliated with the biobank in question.

Another limiting factor was the complexity of the diagnosis of RA. Since there are only classification criteria, which were primarily developed to enable clinical studies to have uniform cohorts for research, and no diagnostic criteria for RA, a diagnosis is ultimately an opinion of the rheumatologist (8, 9, 22). This opinion is based on a subjective combination of clinical signs and symptoms, available clinical tests, differential diagnostics, and knowledge about the epidemiology of the rheumatologist's geographical area (22). Because of this, in our study, fulfilment of the ACR 1987 or ACR/EULAR 2010 classification criteria was not required; however, 88% of the patients with seronegative RA and 97% of the patients with seropositive RA fulfilled one or both of these criteria (8, 9). In our study, a reviewer made the final decision on the correctness of the diagnosis, and a shared decision by two investigators may have strengthened these results. On the other hand, the multicentre design of our study strengthened the external validity of the results.

Conclusion

In summary, the validity of RA diagnoses in Finnish biobank patients was good, especially in patients with entitlement to special reimbursement for medication, more than one visit with the RA diagnosis, and available serological data. In patients with seronegative RA, the validity of a diagnosis of a single visit was only moderate, which is compatible with the notion that seronegative RA is a heterogeneous disease entity. When planning for future studies, it is essential to know the limitations of healthcare registers and the means to manage these limitations. The validity of the RA diagnoses in biobanks can be markedly improved by combining data from different healthcare registers.

Acknowledgements

The study benefited from data from the following biobanks: Auria Biobank, Turku, Finland; Central Finland Biobank, Jyväskylä, Finland; Biobank of Eastern Finland, Kuopio, Finland; Finnish Clinical Biobank Tampere, Tampere, Finland; and Biobank Borealis, Oulu, Finland. Robert M Badeau, MSc, PhD, of Aura Professional English Consulting Ltd (www.auraenglish.com), provided the English language services.

This work was supported by research grants from the FinnGen research project to JP, H-KH, and AP; the Finnish Foundation for Rheumatic Diseases to JP; Turunmaa Duodecim Society to JP; Finska Läkaresällskapet and Stockmanns Foundation to KKE; and the Academy of Finland [grant number 321356] to ASH.

Disclosure statement

No potential conflict of interest was reported by the authors.

Data availability

Due to Finnish national data protection legislation, the register data used in this study cannot be shared without permission from the Health and Social Data Permit Authority of Finland.

References

1. Paskal W, Paskal AM, Debski T, Gryziak M, Jaworowski J. Aspects of modern biobank activity – comprehensive review. Pathol Oncol Res 2018;24:771–85.
2. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK biobank resource with deep phenotyping and genomic data. Nature 2018;562:203–9.
3. Li R, Chen Y, Ritchie MD, Moore JH. Electronic health records and polygenic risk scores for predicting disease risk. Nat Rev Genet 2020;21:493–502.
4. Sund R. Utilisation of administrative registers using scientific knowledge discovery. Intell Data Anal 2003;7:501–19.
5. Sund R. Quality of the Finnish Hospital Discharge Register: a systematic review. Scand J Public Health 2012;40:505–15.
6. Vuori MA, Laukkanen JA, Pietilä A, Havulinna AS, Kähönen M, Salomaa V, et al. The validity of heart failure diagnoses in the Finnish Hospital Discharge Register. Scand J Public Health 2020;48:20–8.
7. Haverinen S, Vihervaara A, Löyttyniemi E, Peltonen S, Koulu L, Tasanen K, et al. Validation of psoriasis diagnoses recorded in Finnish biobanks. Acta Derm Venereol. 2020;100: adv00297doi:10.2340/00015555-3656.
8. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.
9. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2010;69:1580–8.
10. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377–81.
11. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software partners. J Biomed Inform 2019;95:103208.

12. Waldenlind K, Eriksson JK, Grewin B, Askling J. Validation of the rheumatoid arthritis diagnosis in the Swedish National patient register: a cohort study from Stockholm County. BMC Musculoskelet Disord 2014;15:432.
13. Ibfelt EH, Sørensen J, Jensen DV, Dreyer L, Schiøttz-Christensen B, Thygesen PH, et al. Validity and completeness of rheumatoid arthritis diagnoses in the nationwide DanBiO clinical register and the Danish national Patient registry. Clin Epidemiol 2017;9:627–32.
14. Singh J, Holmgren A, Noorbaloochi S. Accuracy of veterans administration databases for a diagnosis of rheumatoid arthritis. Arthritis Rheum 2004;51:952–7.
15. Paalanen K, Puolakka K, Nikiphorou E, Hannonen P, Sokka T. Is seronegative rheumatoid arthritis true rheumatoid arthritis? A nationwide cohort study. Rheumatology 2021;60:2391–5.
16. Paalanen K, Rannio K, Rannio T, Asikainen J, Hannonen P, Sokka T. Does early seronegative arthritis develop into rheumatoid arthritis? A 10-year observational study. Clin Exp Rheumatol 2019;37:37–43.
17. Carrara G, Scirè CA, Zambon A, Cimmino MA, Cerra C, Caprioli M, et al. A validation study of a new classification algorithm to identify rheumatoid arthritis using administrative health databases: case–control and cohort diagnostic accuracy studies. results from the record linkage on rheumatic diseases study of the Italian society for rheumatology. BMJ Open 2015;5:e006029doi:10.1136/bmjopen-2014.
18. Thomas SL, Edwards CJ, Smeeth L, Cooper C, Hall AJ. How accurate are diagnoses for rheumatoid arthritis and juvenile idiopathic arthritis in the general practice research database? Arthritis Rheum 2008;59:1314–21.
19. Widdifield J, Bernatsky S, Paterson JM, Tu K, Ng R, Thorne JC, et al. Accuracy of Canadian health administrative databases in identifying patients with rheumatoid arthritis: a validation study using the medical records of rheumatologists. Arthritis Care Res 2013;65:1582–91.
20. Liao KP, Caim T, Gainer V, Goryachev S, Zeng-Treitler Q, Raychaudhuri S, et al. Electronic medical records for discovery research in rheumatoid arthritis. Arthritis Care Res 2010;62:1120–7.
21. Tenstad HB, Nilsson AC, Dellgren CD, Lindegaard HM, Rubin KH, Lillevang ST. Predictive values of anti-cyclic citrullinated peptide antibodies and rheumatoid factor in relation to serological aspects of the ACR/EULAR 2010 classification criteria for rheumatoid arthritis. Scand J Rheumatol 2020;49:18–20.
22. Aggarwal R, Ringold S, Khanna D, Neogi T, Johnson SR, Mioller A, et al. Distinctions between diagnostic and classification criteria? Arthritis Care Res 2015;67:891–7.

Supplementary material

Supplemental data for this article can be accessed here