Supplementary Table S1. Characteristics of the 60 controls with chest CTs compared to the 173 controls without

	Number (%)		
Characteristic	**With CT (N=60)**	**Without CT (N=173)**	**p-value**
Years to Index Date, median (IQR)	1.2 (-2.8,5.3)	2.2 (-1.2,5.7)	0.41
Age in years, mean (SD)	66 (11)	65 (11)	0.44
Female Sex	54 (90)	132 (76)	0.02
RA Duration in years, mean (SD)	21 (11)	20 (12)	0.45
RF Positive	52 (87)	143 (83)	0.47
CCP Positive	49 (82)	141 (82)	0.98
White, non-Hispanic	53 (88)	166 (96)	0.03
Education Less than College	29 (48)	69 (40)	0.25
Joint Erosions	38 (63)	67 (61)	0.78
Rheumatoid Nodules	22 (37)	58 (34)	0.66
Body Mass Index, kg/m²			0.90
<20	3 (5)	13 (8)	
20-<25	20 (33)	57 (33)	
25-<30	22 (37)	65 (38)	
≥30	15 (25)	38 (22)	
Smoking Status			0.42
Never	30 (50)	94 (54)	
Past	29 (48)	71 (41)	
Current	1 (2)	8 (5)	
Smoking Pack-Years			0.33
Never Smoker	30 (50)	96 (55)	
1-9	14 (23)	31 (18)	
10-19	10 (17)	18 (10)	

20-29	5 (8)	16 (9)	
≥30	1 (2)	12 (7)	
C-Reactive Protein, mg/L			0.92
Normal (<3)	36 (60)	100 (58)	
Low-Positive (3-<10)	16 (27)	51 (29)	
High-Positive (≥10)	8 (13)	22 (13)	
Disease Activity Score-28-CRP			0.82
Remission (<2.6)	27 (45)	77 (45)	
Low (2.6-<3.2)	12 (20)	27 (16)	
Moderate (3.2-<5.1)	17 (28)	53 (31)	
High (≥5.1)	4 (7)	16 (9)	
MDHAQ Score			0.28
0-<0.2	13 (22)	51 (29)	
0.2<-1.0	38 (63)	89 (51)	
≥1.0	9 (15)	33 (19)	
Biologic DMARD Use			0.26
Never	16 (27)	60 (35)	
Past	10 (17)	17 (10)	
Current	34 (57)	96 (55)	
Methotrexate Use			0.95
Never	6 (10)	18 (10)	
Past	25 (42)	68 (39)	
Current	29 (48)	87 (50)	
Prednisone Use			0.11
Never	6 (10)	25 (14)	
Past	32 (53)	108 (62)	
Current	22 (37)	40 (23)	

CRP = C-reactive protein, CT = computed tomography, DMARD = disease-modifying anti-rheumatic

drug, ILD = interstitial lung disease, IQR = interquartile range, MDHAQ = multi-dimensional health assessment questionnaire, RA = rheumatoid arthritis, RF = rheumatoid factor, SD = standard deviation

STROBE Statement—Checklist of items that should be included in reports of ***case-control studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
Methods			
Study design	4	Present key elements of study design early in the paper	5-6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	6
		(*b*) For matched studies, give matching criteria and the number of controls per case	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	6,7,17
Study size	10	Explain how the study size was arrived at	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7
		(*b*) Describe any methods used to examine subgroups and interactions	7
		(*c*) Explain how missing data were addressed	7
		(*d*) If applicable, explain how matching of cases and controls was addressed	6
		(*e*) Describe any sensitivity analyses	7
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	7
Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure	Table1

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Tables 1-4, 7
		(*b*) Report category boundaries when continuous variables were categorized	Table 1
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12
Discussion			
Key results	18	Summarise key results with reference to study objectives	14
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	18
Generalisability	21	Discuss the generalisability (external validity) of the study results	17
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	18

*Give information separately for cases and controls.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

Analysis Plan: Investigating Risk Factors for RA-ILD

Descriptive Statistics

- N for RA-ILD cases and controls (Note: please use mean for normal variables and median for non-normal ones)
    - Age (matched)
    - Sex (matched)
    - RA duration (matched)
    - RF positivity (matched)
    - CCP positivity (yes or no)
    - Race/ethnicity (white non-Hispanic vs other)
    - Education (lower education vs not)
    - Methotrexate use (never, past, current)
    - Prednisone (never, past, current)
    - DMARD use (never, past, current)
    - Disease activity **(**DAS28-CRP continuous and <2.6, 2.6-<3.2, 3.2-<5.1, ≥5.1)
    - Smoking status (never, past, current)
    - Smoking pack-years (continuous and 0, 1-9, 10-19, 20-29, or 30+)
    - Functional assessment (MD-HAQ, continuous)
    - BMI (continuous and <20, 20-<25, 25-<30, or ≥30 kg/m^2)
    - CRP level (continuous and normal, low-mid positive, high positive)
    - Radiographic changes (yes or no)
    - Rheumatoid nodules (yes or no)
- Unadjusted OR (95% CI) for the above variables

Aim 0: Confirm association of previously-shown risk factors for RA-ILD including CCP positivity, drug use, and disease activity. We hypothesize CCP positivity, drug use, and increased disease activity will continue to be associated with increased risk of RA-ILD even after adjusting for confounders.

- Above models also adjusting for seropositivity (CCP or RF), race/ethnicity, education, BMI (continuous if linear otherwise categorical), and smoking pack-years (continuous if linear otherwise categorical)
    - Race/ethnicity (white non-Hispanic vs other)
    - Education (lower education vs not)
    - CCP positivity (yes or no)
    - Methotrexate use (never, past, current)
    - Prednisone (never, past, current)
    - DMARD use (never, past, current)
    - Disease activity **(**DAS28-CRP continuous and <2.6, 2.6-<3.2, 3.2-<5.1, ≥5.1)

Aim 1: Evaluate whether a threshold effect exists for smoking on risk of RA-ILD. We hypothesize that similar to RA, approximately 20 pack-years of smoking produces the greatest risk increase for RA-ILD.

- Above models also adjusting for seropositivity (CCP or RF), race/ethnicity, education, BMI (continuous if linear otherwise categorical)
    - Smoking status (never, past, current)
    - Smoking pack-years (restricted cubic spline if enough cases; otherwise 0, 1-9, 10-19, 20-29, or 30+)

Aim 2: Investigate the association between novel predictors and RA-ILD including functional status, obesity, CRP level, radiographic changes, and rheumatoid nodules. We hypothesize that functional status, obesity, CRP level, radiographic changes, and rheumatoid nodules are associated with increased risk of RA-ILD.

- Above models also adjusting for seropositivity (CCP or RF), race/ethnicity, education, BMI (continuous if linear otherwise categorical), and smoking pack-years (continuous if linear otherwise categorical)
    - Functional assessment (MD-HAQ, restricted cubic spline if enough cases; otherwise categorize)
    - BMI (<20, 20-<25, 25-<30, or ≥30 kg/m^2)
    - CRP level (normal, low-mid positive, high positive)
    - Radiographic changes (yes or no)
    - Rheumatoid nodules (yes or no)

Aim 3: Determine the degree to which known risk factors for RA-ILD can predict disease development. We hypothesize that RA factors contribute the most to RA-ILD risk and that together, known risk factors for RA-ILD can account for 80% of disease risk (i.e. AUC of 0.8).

- Calculate AUC for demographic factors (age, sex, race/ethnicity, and education)
- Calculate AUC for lifestyle factors (BMI, smoking pack-years)
- Calculate AUC for RA factors (RA duration, RF positivity, CCP positivity, drug use, disease activity, functional assessment, CRP level, radiographic changes, rheumatoid nodules)
- Calculate AUC for demographic, lifestyle, and RA factors together
- Create ROC curve graph for the above four scenarios

Limitations

- AUC is higher (around 0.9) in other papers. It's likely lower here because we made controls more similar to RA-ILD cases by matching on RA duration and RF positivity.
- These analyses don't account for the time-varying nature of the covariates.
- Limitation: timing of ILD can be insidious, so predictors still might be after onset of RA-ILD symptoms. Could look at dyspnea score from BRASS.
-