

Figure S1 The original Chinese version of the rheumatoid arthritis knowledge graph (partially displayed).

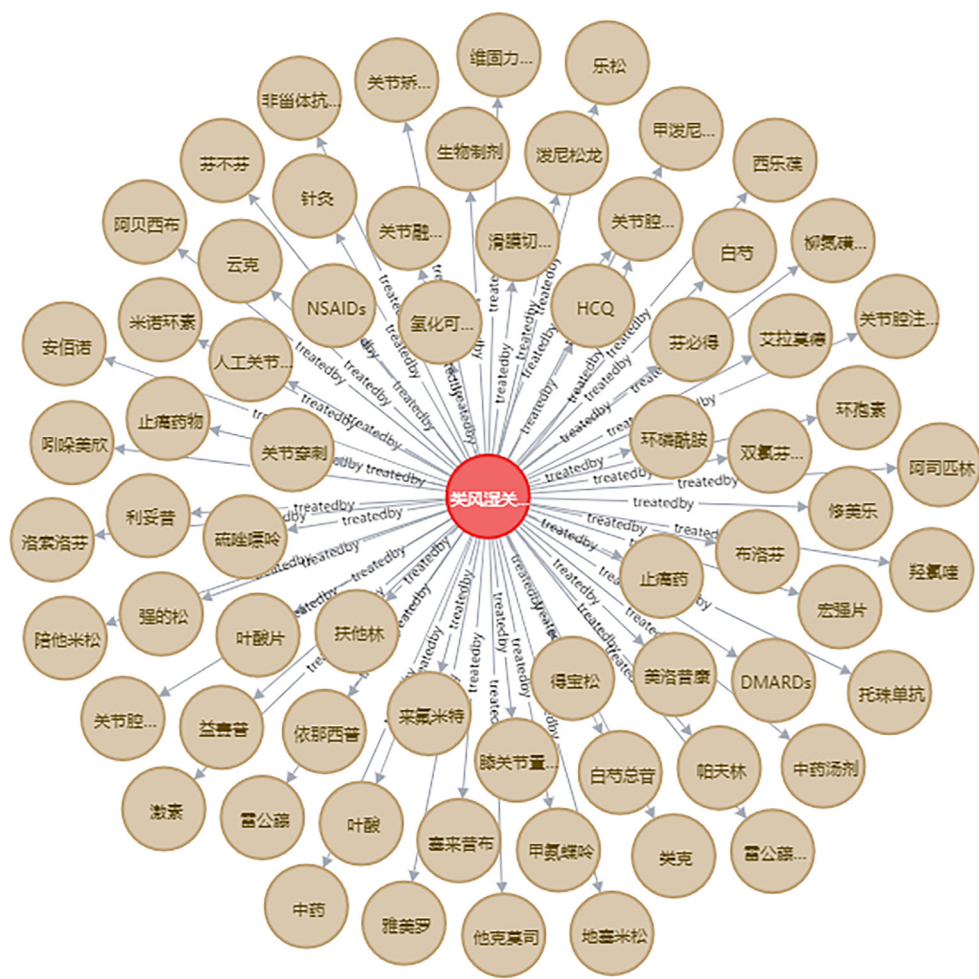


Figure S2 The original Chinese version of the rheumatoid arthritis treatment subgraph.

Table S1 Detailed parameter settings of the NER model

Parameter name	Parameter value
Maximum sentence length	128
Dropout value	0.1
MC-BERT layers	12
MC-BERT hidden layers	768
MC-BERT learning rate	2e-5
CRF learning rate	5e-5
LSTM learning rate	1e-3

NER, named entity recognition; MC, medical; BERT, bidirectional encoder representation from transformers; LSTM, long short-term memory; CRF, conditional random field.

Table S2 Detailed parameter settings of the relation extraction model

Parameter name	Parameter value
Maximum sentence length	128
Dropout value	0.1
MC-BERT layers	12
MC-BERT hidden layers	768
MC-BERT learning rate	2e-5
FFNN hidden layers	768

MC, medical; BERT, bidirectional encoder representation from transformers; FFNN, feedforward neural network.

Table S3 Entity type and number of 367 manually labeled CEMRs

Entity category	Number
Disease	1,435
Symptom	7,812
Treatment	8,097
Physical examination	3,094
Imaging test	1,358
Lab test	7,675
Aggravating factor	68
Mitigation factor	206
Body part	10,518

CEMRs, Chinese electrical medical records.

Table S4 Entity relation type and number of 367 manually labeled CEMRs

Relation category	Number
Disease and symptom	4,761
Disease and physical examination	4,305
Disease and lab test	23,146
Disease and imaging test	4,325
Disease and treatment	5,634
Treatment improved symptom	862
Treatment did not improve symptom	605
Body part and symptom	10,950
Body part and physical examination	4,003
Aggravating factor and symptom	91
Mitigation factor and symptom	224
All	58,906

CEMRs, Chinese electrical medical records.