

An automated approach to extract pharmacokinetic parameters from scientific publications

Frank Kloprogge
Pharmacokinetics-Pharmacodynamics Group
UCL Institute for Global Health

Acknowledgements



▶ UCL team

- ▶ Ferran Gonzalez Hernandez
- ▶ Victoria Smith
- ▶ Waty Lilaonitkul
- ▶ Frank Klopogge
- ▶ Joseph Standing

▶ Contributors and Collaborations

- ▶ Juha Iso-Sipila
- ▶ Paul Goldsmith
- ▶ James Yates
- ▶ Ahmed Allousa
- ▶ Simon Carter
- ▶ José Antonio Cordero Rigol
- ▶ Maria Rosa Ballester
- ▶ Mario Duran Hortolà
- ▶ Albert Solé Guixeres
- ▶ Palang Chotsiri
- ▶ Thanaporn Wattanakul
- ▶ Gill Mundin

Rational

- 1. PK predictions for early drug development**
- 2. Development of population pharmacokinetic models**
- 3. Pop PK models for TDM and personalised medicine**

Model pipeline

1. Identify PK papers
2. NER & REX
3. Table extraction

Identify PK papers

1. Search term: "Pharmacokinetics"
2. Training corpus
 - ▶ Two independent labelers
 - ▶ 3,992 articles
3. Test corpus
 - ▶ Three independent labelers
 - ▶ 800 articles

Field	Training	Final test
Title	100	100
Abstract	87.17	87.67
Authors	99.44	99.63
Journal	100	100
Publication Type	100	100
Keywords	15.41	16.125
MeSH terms	97.67	98.25
Chemicals	93.86	94.13
Affiliations	80.94	79.125
Label		
Relevant	19.81	20.25
Not Relevant	80.19	79.75

Identify PK papers

- ▶ XGBoost
- ▶ Unigrams
- ▶ Biobert mean pooling
- ▶ 200 bootstraps
 - ▶ Training 60%
 - ▶ Development 20%
 - ▶ Test 20%

▶ Training data

Pipeline	Precision (%)	Recall(%)	F_1 (%)	F_1 IQV
Unigr.	80.1 (73.9,86.0)	82.3 (74.1,88.6)	80.6 (75.8,85.2)	9.4
Unigr. + mean pool	83.7 (76.7,89.1)	80.4 (74.1,87.3)	81.7 (77.8,86.0)	8.2
Unigr. + mean + min&max	83.8 (75.6,88.8)	79.1 (73.4,85.4)	81.0 (77.2,85.4)	8.2

▶ Test data

Precision (%)	Recall(%)	F_1 (%)	Accuracy (%)
84.8%	82.8%	83.8%	93.2%

Gonzalez Hernandez F, Carter SJ, Iso-Sipilä J, Goldsmith P, Almousa AA, Gastine S, Lilaonitkul W, Klopogge F, Standing JF. An automated approach to identify scientific publications reporting pharmacokinetic parameters. Wellcome Open Res. 2021 Apr 21;6:88. doi: 10.12688/wellcomeopenres.16718.1. PMID: 34381873; PMCID: PMC8343403.

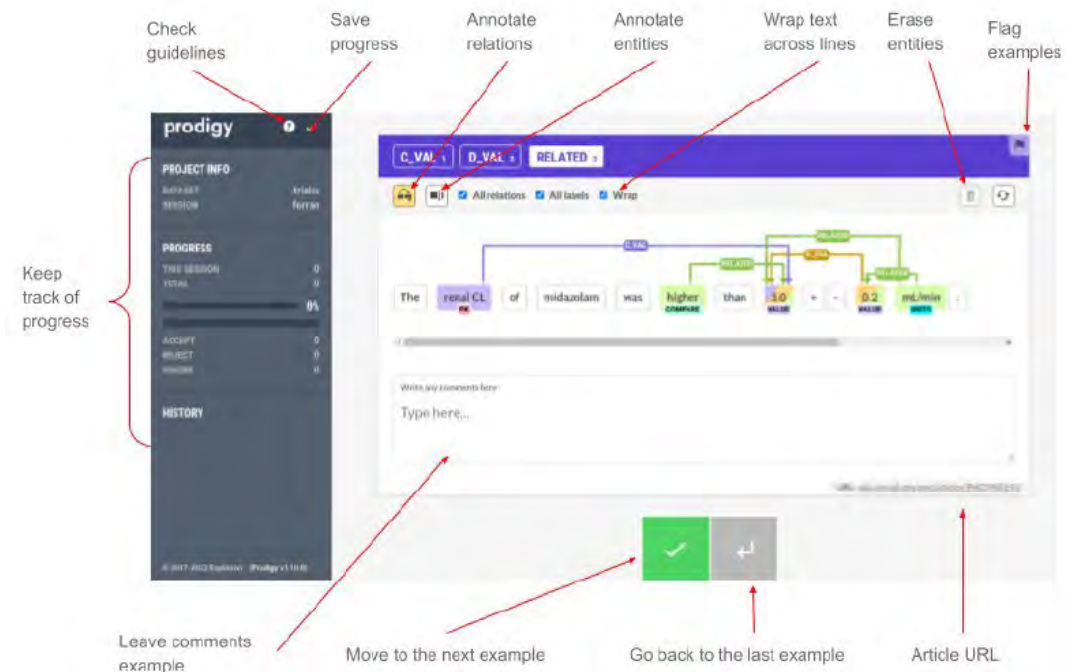
NER & REX

NER



Dataset	# Sentences	# entity mentions	% sentences with PK mentions	% of full-text sentences
Training	2800	3680	64.25	79.46
Development	500	149	16.40	50.8
Test	1500	390	16.40	50.8

REX

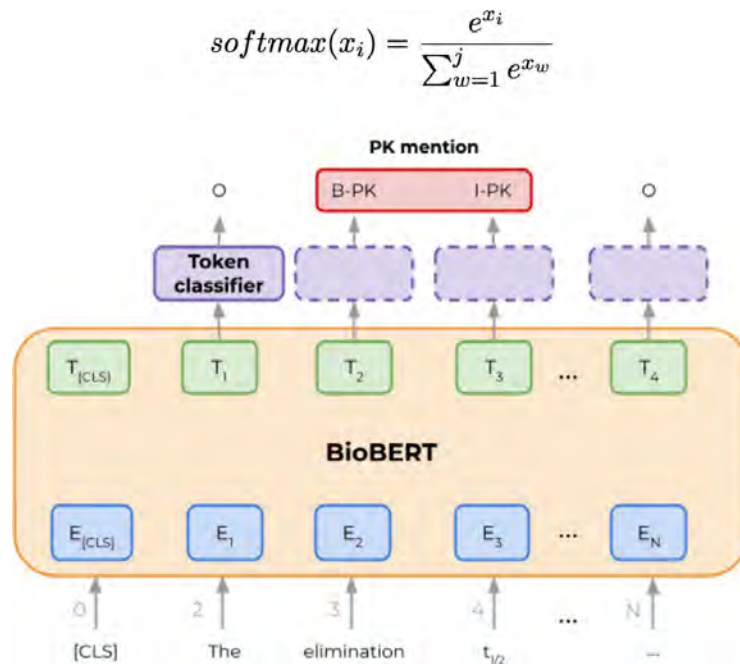


		Training	Development	Test	Total
Sentences	Amount #	2100	500	1000	3600
	with relations (%)	57.05	53.00	56.80	56.42 [†]
	from full-text (%)	48.71	50.00	50.30	49.33 [†]
Entities	PK	1890	394	856	3140
	Units	2286	474	1056	3816
	Value	3524	702	1557	5783
	Range	314	74	174	562
	Compare	51	18	34	103
Relations	<i>Central_{val}</i>	2794	571	1312	4677
	<i>Deviation_{val}</i>	1049	207	419	1675
	<i>Related</i>	3643	764	1652	6059

NER & REX

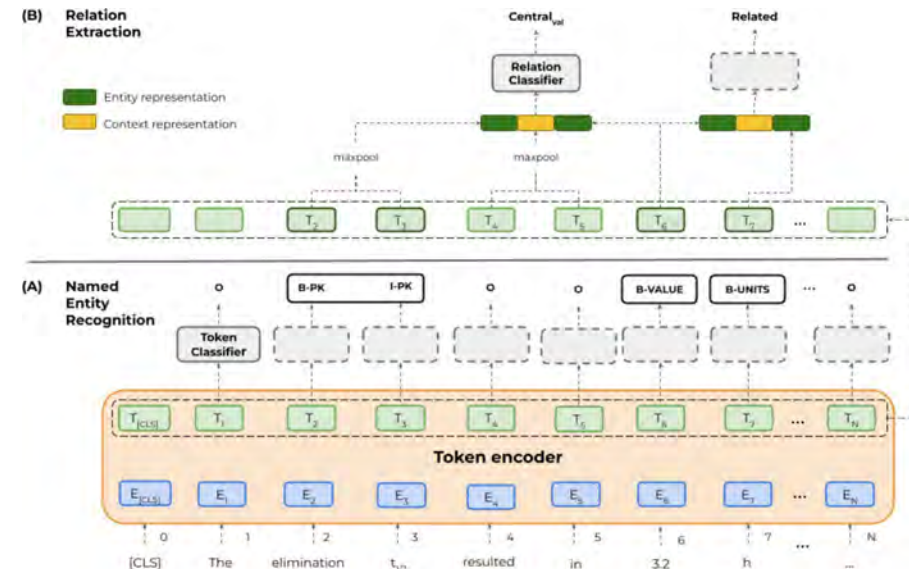
NER

- ▶ Neural network
 - ▶ Spacy (Convolutional Neural Network)
 - ▶ Best performing BioBERT (pre-training with PubMed abstracts)



REX

- ▶ Multi-task learning
 - ▶ Architecture the same
 - ▶ NER for PK, units etc simultaneously
 - ▶ Tweaked to do relation extraction simultaneously



NER & REX

NER

Model	Strict			Partial		
	P	R	F_1	P	R	F_1
Rule-based	52.8	43.59	47.75	69.25	57.18	62.64
ScispaCy	77.09	82.82	79.85	80.91	86.92	83.81
BERT	81.47	87.72	84.48	84.92	91.43	88.05
BioBERT	90.49	90.26	90.37	92.54	92.31	92.43

REX

Entity	Precision		Recall		F1	
	MT	no-MT	MT	no-MT	MT	no-MT
PK	90.82 _{4.02}	89.98 _{3.86}	90.57 _{3.76}	90.09 _{3.05}	90.39 _{2.1}	90.02 _{1.72}
Units	95.49 _{1.87}	95.79 _{1.66}	96.17 _{2.07}	95.69 _{3.85}	95.65 _{0.68}	95.56 _{1.52}
Value	94.83 _{2.78}	94.96 _{2.87}	96.18 _{3.17}	95.21 _{5.94}	95.54 _{2.53}	95.04 _{2.02}
Range	93.49 _{4.9}	93.28 _{6.24}	90.26 _{8.22}	87.39 _{10.33}	91.66 _{4.41}	90.4 _{3.71}
Compare	88.23 _{6.81}	88.23 _{16.99}	66.67 _{9.09}	68.18 _{11.44}	76.53 _{5.82}	75.64 _{8.12}
Micro-average					94.03 _{1.63}	93.69 _{1.60}
Macro-average					90.02 _{2.23}	89.56 _{2.45}

Initial papers #	Papers with abstract # (%)	Abstracts with PK estimates # (%)	PK estimates #
120,913	113,415 (93.80%)	65,377 (57.64%)	274,277

Home Paper Search *NER demo* *REX demo* PKDB Team Contact

Find estimates of PK parameters from scientific abstracts

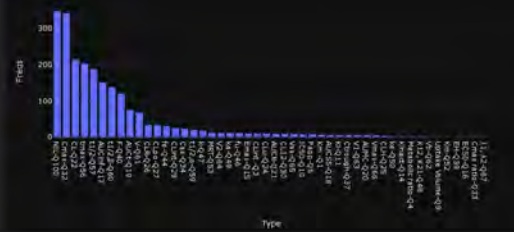
An app to collate estimated pharmacokinetic (PK) parameters *in vivo* from scientific abstracts.
Note: This is still a demo app which is currently under development.

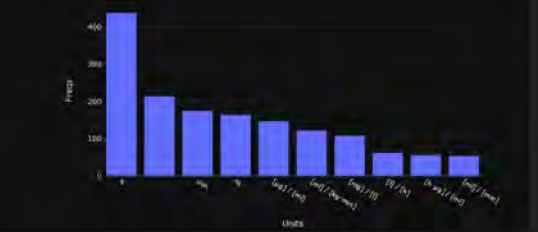
How does it work?

paracetamol Search

ID	Parameter	Type	Value	Units	Compare	Sentence	Title	URL	lit
2610379	formation clearance	CLint-Q3	1.45	[ml] / [kg-min]		The formation clearance of paracetamol in the rat is 1.45 ml/kg/min.			
2610379	formation clearance	CLint-Q3	0.264	[ml] / [kg-min]		The formation clearance of paracetamol in the rat is 0.264 ml/kg/min.			
2610379	renal excretion rate constants	ka-Q30	0.388	[ml] / [kg-min]		The renal excretion rate constants of paracetamol in the rat is 0.388 ml/kg/min.			
2751775A	plasma tmax	tmax-Q36	195	min		The median plasma Tmax of paracetamol in the rat is 195 min.			
2751775A	plasma tmax	tmax-Q36	75	min		The median plasma Tmax of paracetamol in the rat is 75 min.			
10091939	hepatic clearance	CLH-Q25	10.4	[ml] / [min]		Mean +/- S. D. hepatic clearance of paracetamol in the rat is 10.4 ml/min.			
10091939	apparent intrinsic clearance	CLint-Q3	94.1	[ml] / [min]		Mean +/- S. D. hepatic clearance of paracetamol in the rat is 94.1 ml/min.			
10091939	hepatic extraction ratio	ER-Q39	04.1	[ml] / [min]		Mean +/- S. D. hepatic extraction ratio of paracetamol in the rat is 04.1 ml/min.			
10091939	renal excretion half life	t1/2-Q38	88.9	min		Mean +/- S. D. hepatic extraction ratio of paracetamol in the rat is 88.9 min.			
114883	excretion rates	ra-Q44	0.131 to 0.345	[mg] / [d+kg]		Twelve healthy 2- to 3-month-old rats were given paracetamol at a dose of 100 mg/kg.			

Abstracts: 467
 # Estimates: 2217

Entity frequency: 

Top Units: 

Case study

ORIGINAL ARTICLE

Scaling beta-lactam antimicrobial pharmacokinetics from early life to old age

Correspondence Dr Dagan O. Lonsdale, Institute for Infection and Immunity, Room 1.141 St George's, University of London, London, SW17 0RE UK. Tel.: +44 208 725 0205; E-mail: dlonsdale@sgul.ac.uk

Received 2 April 2018; Revised 2 August 2018; Accepted 22 August 2018

Dagan O. Lonsdale^{1,2}, Emma H. Baker^{1,2}, Karin Kipper^{1,3,4}, Charlotte Barker¹, Barbara Philips^{1,2}, Andrew Rhodes², Mike Sharland^{1,2} and Joseph F. Standing^{1,2,5,6}

Presence of parameters in abstract

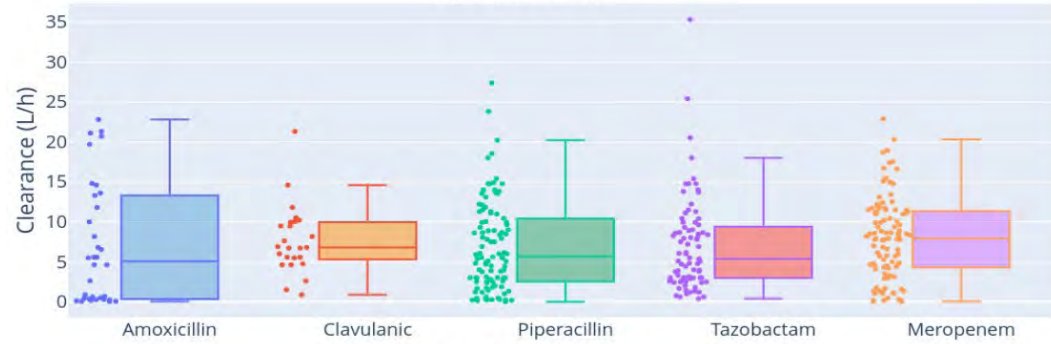
1. Amoxicillin ± clavulanate
 - ▶ Clearance 43.48 %
 - ▶ Volume of distribution 43.48 %
2. Piperacillin-tazobactam
 - ▶ Clearance 42.49 %
 - ▶ Volume of distribution 33.33 %
3. Meropenem
 - ▶ Clearance 45.28 %
 - ▶ Volume of distribution 41.51 %

Recall rates

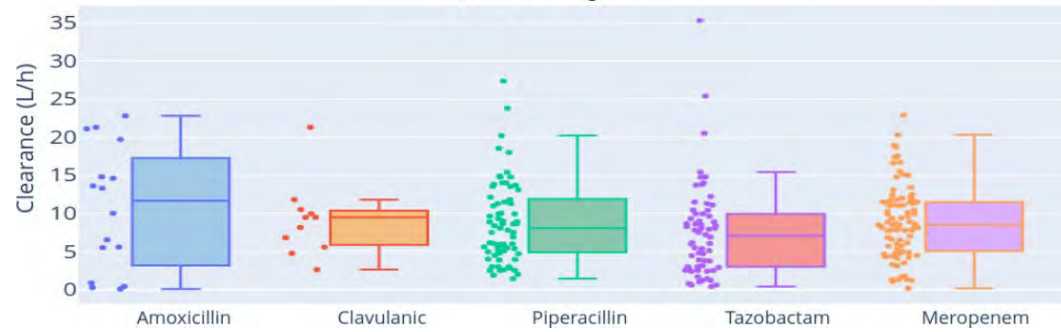
1. Amoxicillin ± clavulanate
 - ▶ Clearance 100 %
 - ▶ Volume of distribution 91.67 %
2. Piperacillin-tazobactam
 - ▶ Clearance 97.05 %
 - ▶ Volume of distribution 87.50 %
3. Meropenem
 - ▶ Clearance 91.67 %
 - ▶ Volume of distribution 94.45 %

Case study

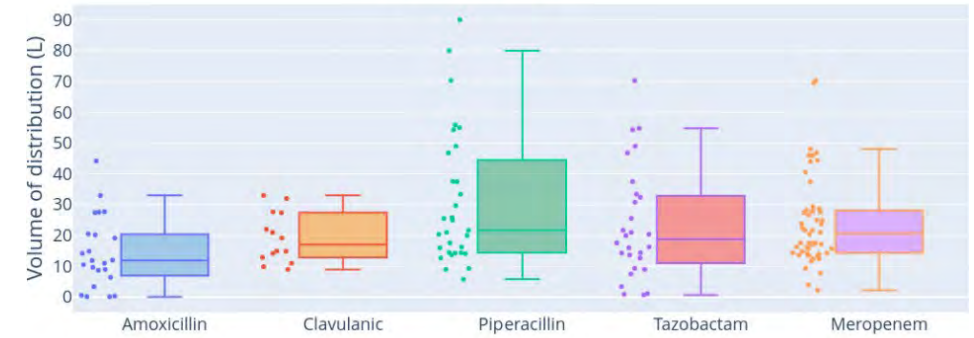
A) Before cleaning



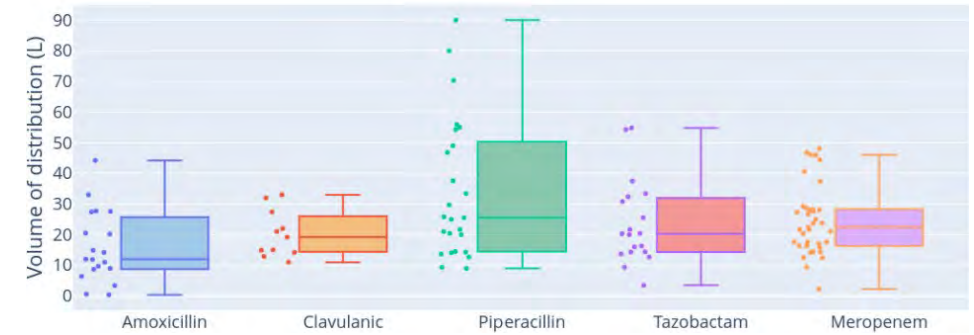
B) After cleaning



A) Before cleaning



B) After cleaning



Work in progress: Full text machine reading

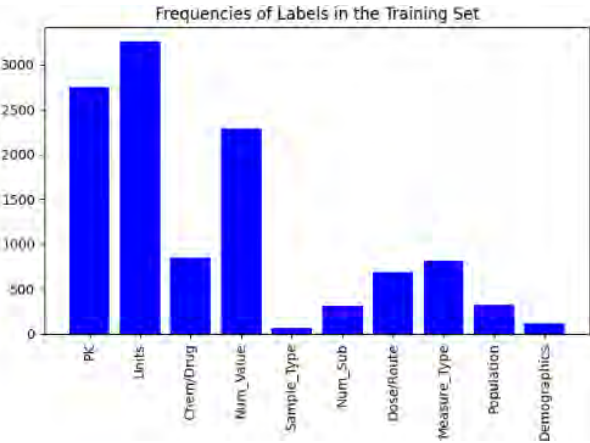
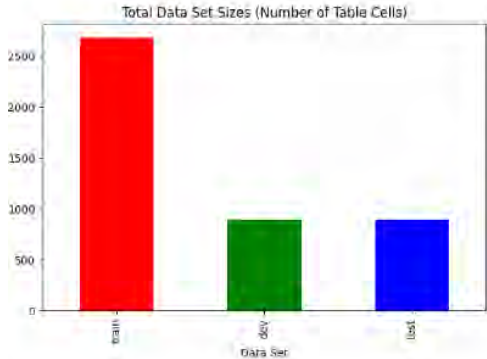
- ▶ **Parsed XML PK papers:**

- ▶ Total Number of XML PK Papers: 7,353
- ▶ Total Number of Sentences in the PK Papers (excluding the abstract): 1,159,341
- ▶ Total Number of Results Sentences: 263,092
- ▶ Total Number of Numeric Sentences in Results Sections: 200,880

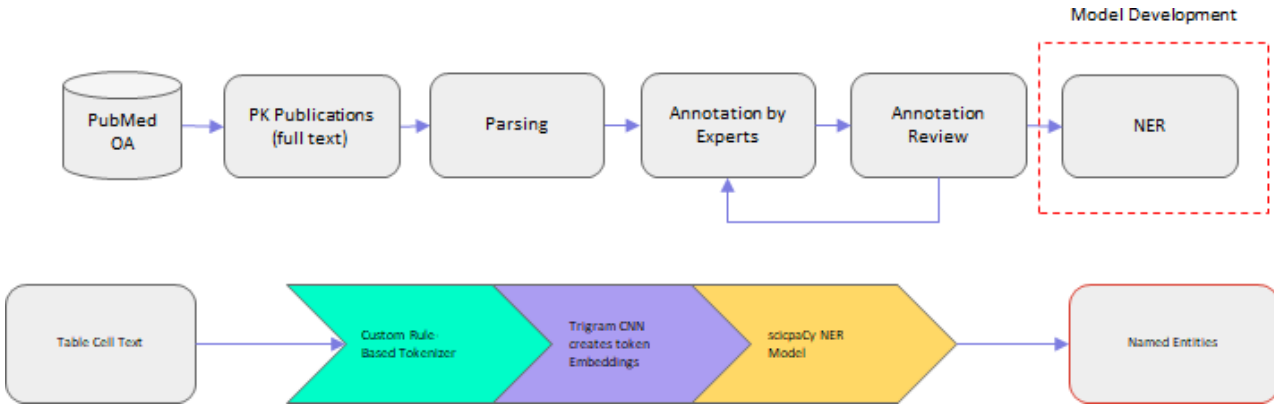
- ▶ **Machine reading ongoing**

Work in progress: Table extraction

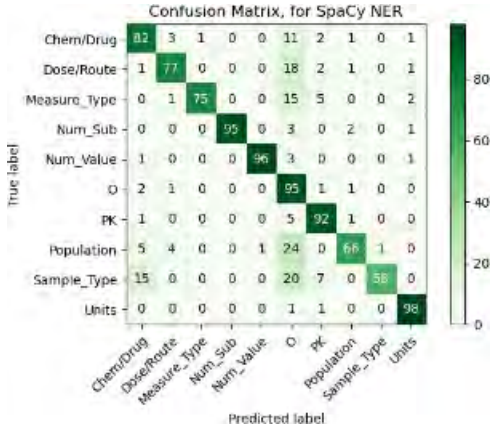
	0	1	2	3	4
0	Parameters	Results (median [interquartile range]) for:	Results (median [interquartile range]) for:	Results (median [interquartile range]) for:	Results (median [interquartile range]) for:
1	Parameters	Risperidone (n = 43)	loperamide (n = 46)	Rizatriptane (n = 36)	Ethinamate (n = 47)
2	Cross (log/ml)	4.13 (2.47-5.60)	3.970 (2.979-4.544)	34.21 (3000-41.80)	2.278 (1.894-3.10)
3	Time (hr)	3.00 (2.00-4.00)	2.00 (1.00-3.00)	2.00 (1.00-3.00)	3.00 (2.00-4.00)
4	AUCD (last (log/ml-hr))	16.42 (11.97-24.28)	21.80 (13.80-33.89)	275.10 (173.7-388.5)	16.72 (12.72-22.93)
5	AUCD (0 (log/ml-hr))	21.32 (13.57-28.60)	22.50 (14.75-34.59)	386.6 (320.0-463.9)	20.41 (16.15-26.27)
6	Cl (2/h)	2.45 (1.86-3.08)	3.92 (3.18-4.72)	4821 (3.71-6.04)	7.507 (6.117-8.69)
7	Cl/F (8000/h)	25.11 (16.62-39.41)	11.60 (7.179-18.42)	3.977 (2.66-4.78)	48.39 (34.31-62.73)
8	Vo/F (liter)	80.95 (67.21-113.70)	66.67 (48.25-92.37)	25.03 (28.24-42.51)	488.40 (376.90-630.70)
9	Vo/F (liter)	130.20 (90.75-202.40)	72.35 (57.56-98.33)	38.74 (31.44-48.13)	498.90 (410.60-722.88)



Work in progress: Table extraction



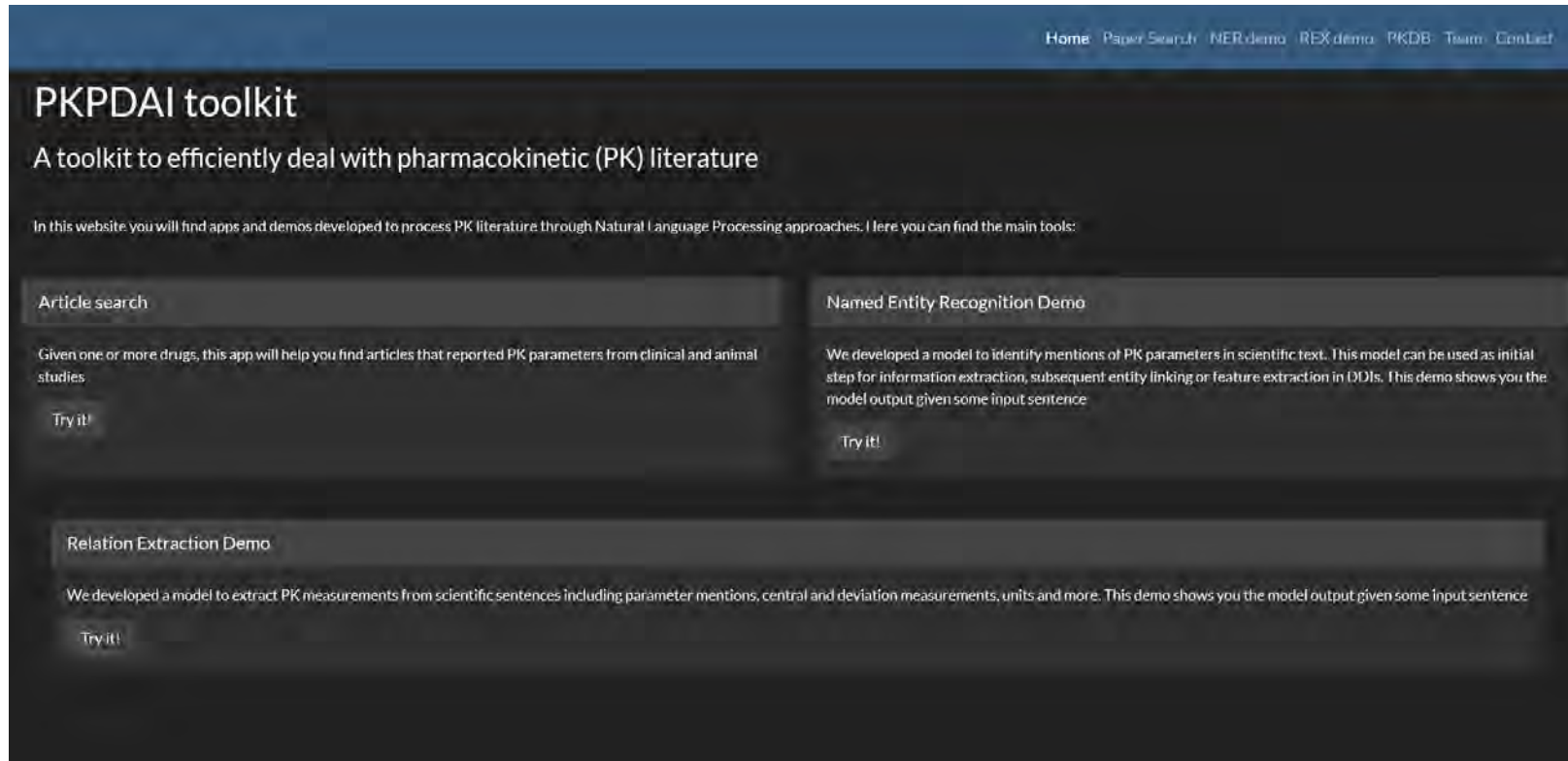
► Confusion matrix



► Performance

Entity Label	Annotator F1-score	Model F1-score
Units	0.95	0.98
Numeric Value	0.95	0.97
PK Parameter	0.97	0.96
Number of Subjects	0.95	0.94
Measure Type	0.90	0.94
Chem/Drug	0.94	0.92
Dose/Route	0.82	0.85
Sample Type	0.87	0.83
Population	0.74	0.81
Macro-F1	0.90	0.91

Please use PKPDai!



The screenshot shows the PKPDai toolkit website. At the top right, there is a navigation menu with links for Home, Paper Search, NER demo, REX demo, PKDB, Team, and Contact. The main heading is "PKPDai toolkit" followed by the subtitle "A toolkit to efficiently deal with pharmacokinetic (PK) literature". Below this, a paragraph states: "In this website you will find apps and demos developed to process PK literature through Natural Language Processing approaches. Here you can find the main tools:". There are three tool cards: 1. "Article search" with the description "Given one or more drugs, this app will help you find articles that reported PK parameters from clinical and animal studies" and a "Try it!" button. 2. "Named Entity Recognition Demo" with the description "We developed a model to identify mentions of PK parameters in scientific text. This model can be used as initial step for information extraction, subsequent entity linking or feature extraction in DDIs. This demo shows you the model output given some input sentence" and a "Try it!" button. 3. "Relation Extraction Demo" with the description "We developed a model to extract PK measurements from scientific sentences including parameter mentions, central and deviation measurements, units and more. This demo shows you the model output given some input sentence" and a "Try it!" button.

info@pkpdai.com | <https://pkpdai.com/>

Backup-slides

