Large Language Models and Case Reports: A New Approach for Real-World Data in Rare Disease Natural History Analysis

RWD51

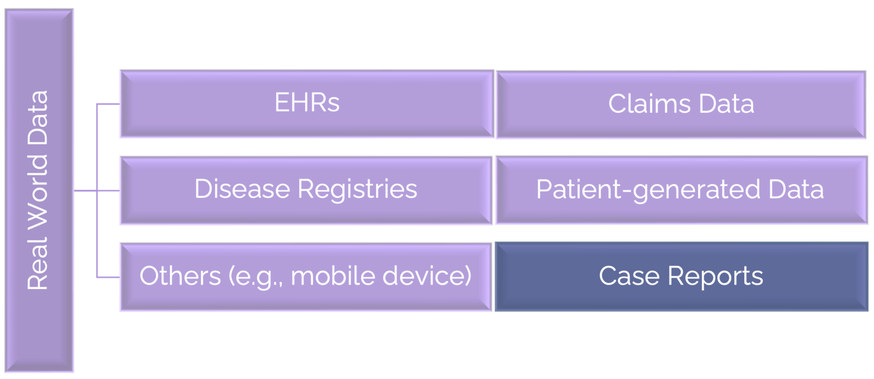
H. Paek, PhD[1*], K. Lee, PhD[1*], LC. Huang, PhD[1], A. Annan, PhD[1], X. Wang, PhD[1]

[1]IMO Health, Rosemont, IL, USA

[*]Equal Contribution

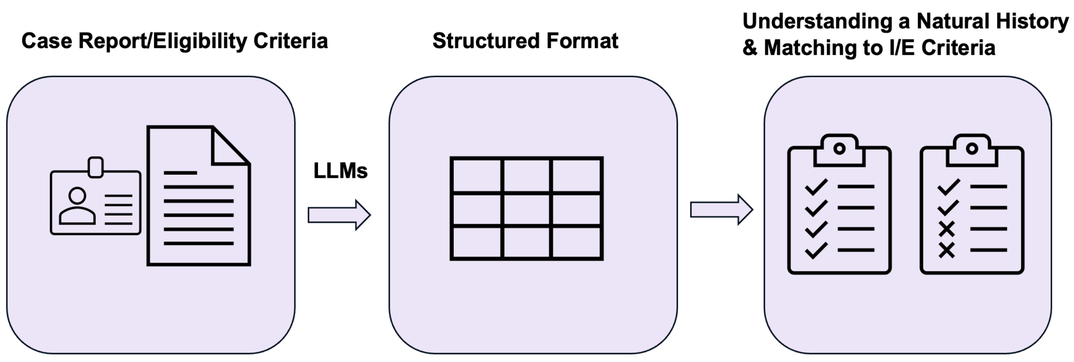
OBJECTIVE

- Exploring Case Reports as Potential Real-World Data For the Natural History and External Control Arms of Rare Disease like IgA Nephropathy and Fabry Disease



METHODS

- Converting Unstructured Case Reports and Clinical Trial Eligibility Criteria into Structured Data Using Large Language Models (LMs)[1,2]



RESULTS

1. LLM-based Model Performance

	Precision	Recall	F 1 Score	Average No. of Data Elements (concepts and values) Compared
Fabry Disease	99.56%	99.66%	99.61	286 (ranging from 129 to 452)
IgA Nephropathy	98.35%	97.36%	97.85	94 (ranging from 67 to 127)

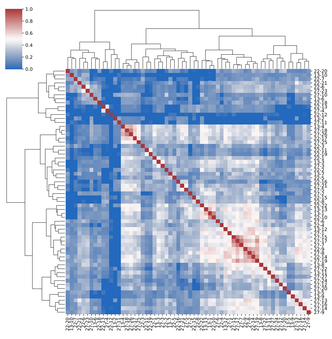
2. Cohort Demographics and Clinical Features extracted

	IgAN (N=83)		Fabry Disease (N=59)	
Characteristics	Male (%)	Female (%)	Male (%)	Female (%)
Age				
Mean (SD)	41.2 (20.8)	42.9 (17.8)	42.1 (18.9)	54.4 (12.0)
Race and Ethnicity				
Caucasian	4 (7.8)	1 (3.1)	6 (14.6)	2 (11.1)
Chinese	5 (9.8)	4 (12.5)	2 (4.9)	0 (0.0)
Japanese	5 (9.8)	2 (6.3)	3 (7.3)	2 (11.1)
Others	7 (13.7)	1 (3.1)	1 (2.4)	1 (5.6)
Not Specified	30 (58.8)	24 (75.0)	29 (70.7)	13 (72.2)
Total	51	32	41	18

Clinical Features Extracted
Patient Demographics
Genetic Information (Diagnosis)
Lab Test Results
Treatment History
Comorbidities
Etc..

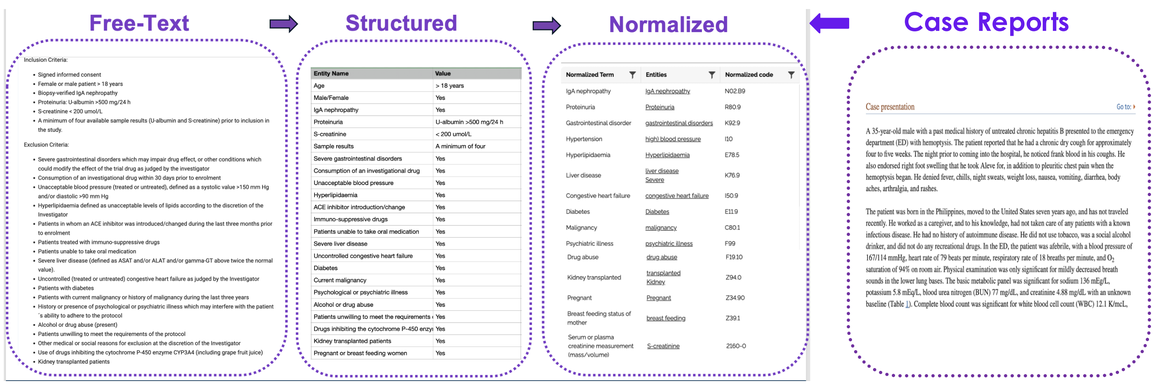
3. Similarity Analysis based on Clinical Features: LLM Explanation

Age	Date	Event Name	Event Values	Comments
32	Not given	Admission to hospital	Generalized edema for 1 month, BP: 145/96 mmHg, scattered angiokeratoma, ed	Initial presentation
32	Not given	Laboratory work-up	Urinalysis: 3+ protein, Urinary total protein: 6.38 g/d, Serum albumin: 28 g/L, Seru	Initial lab results
32	Not given	Electrocardiogram	Sinus tachycardia, high P waves in II, III, aVF leads, left ventricular high voltage	Initial cardiac assessment
32	Not given	Kidney Ultrasonography	Elevated echogenicity of parenchyma, no obvious kidney shrinkage	Initial kidney imaging
32	Not given	Other Examinations	Transthoracic echocardiography, audiometry, ophthalmology, brain MRI: Norma	Initial comprehensive assessment
32	Not given	Diagnosis	Nephrotic Syndrome (NS)	Initial diagnosis
32	Not given	Treatment Initiation	Methylprednisolone: 32 mg/day, Tacrolimus: 1.5 mg twice/day, Valsartan	Initial treatment plan
32	Not given	Renal Biopsy and Pathologi	Light microscopy: vacuolization in podocyte cytoplasm, mild chronic and acute t	Renal pathology findings
32	Not given	Genetic Testing	GLA gene: c.707G > A (p.Trp236*), PKD1 gene: c.988A > G (p.Asn329Ser)	Genetic findings
32	Not given	α-Gal A Activity Testing	Proband: 0.31 nmol/h/mg (normal: 31.30–51.00 nmol/h/mg)	Enzyme activity testing
32	8 weeks	Treatment Adjustment	Methylprednisolone reduced to 24 mg/day, 24-h proteinuria: ~1.0 g	Response to initial treatment
32	3 months	Treatment Adjustment	Tacrolimus discontinued	Adjustment in treatment
32	6 months	Treatment Adjustment	Methylprednisolone tapered to 8 mg/day	Continued tapering of steroids
33	1 year	Treatment Maintenance	Methylprednisolone maintained at 4 mg/day	Long-term maintenance dose
33	Follow-up	Follow-up Results	Proteinuria: ~0.5 g/d, Serum albumin: 44 g/L, No other organ involvement	Follow-up outcomes
32	Not given	Enzyme Replacement Thera	Not administered due to financial constraints	Financial constraints prevented ERT

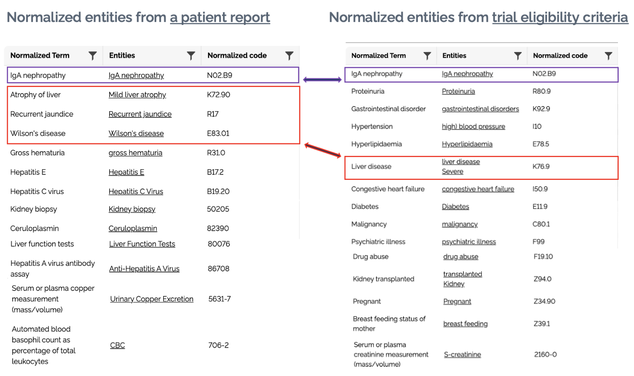


RESULTS

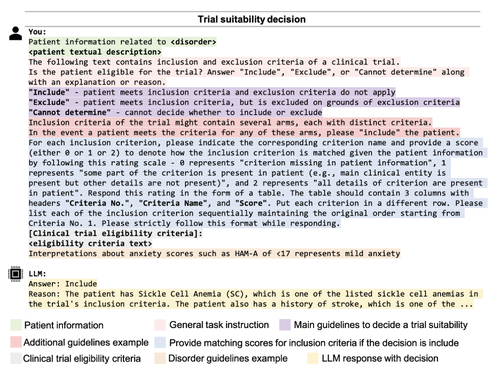
1. Structured & Normalized Criteria Eligibility Criteria & Case reports



2. Normalized Patient-Trial Information Matching



3. LLM-Assisted Patient-Trial Matching[3]



CONCLUSIONS

- **Our study validates case reports as valuable sources of RWD and demonstrates the effectiveness of LLM in extracting clinical information from case reports.**
- **This approach enhances our understanding of the natural history of rare diseases and demonstrates potential source for external.**

References

1. AutoCriteria: J Am Med Assoc. 2024 31 (2): 375-385
2. SEETrials: Inform Med Unlocked 2024 50: 101589
3. Patient2Trials: Inform Med Unlocked Under Revision

CONTACT INFORMATION

klee@imohealth.com
hpaek@imohealth.com
xwang@imohealth.com