

ORIGINAL RESEARCH ARTICLE

Factors associated with social determinants of health mentions in PubMed clinical case reports from 1975 to 2022: A natural language processing analysis

Supplementary File

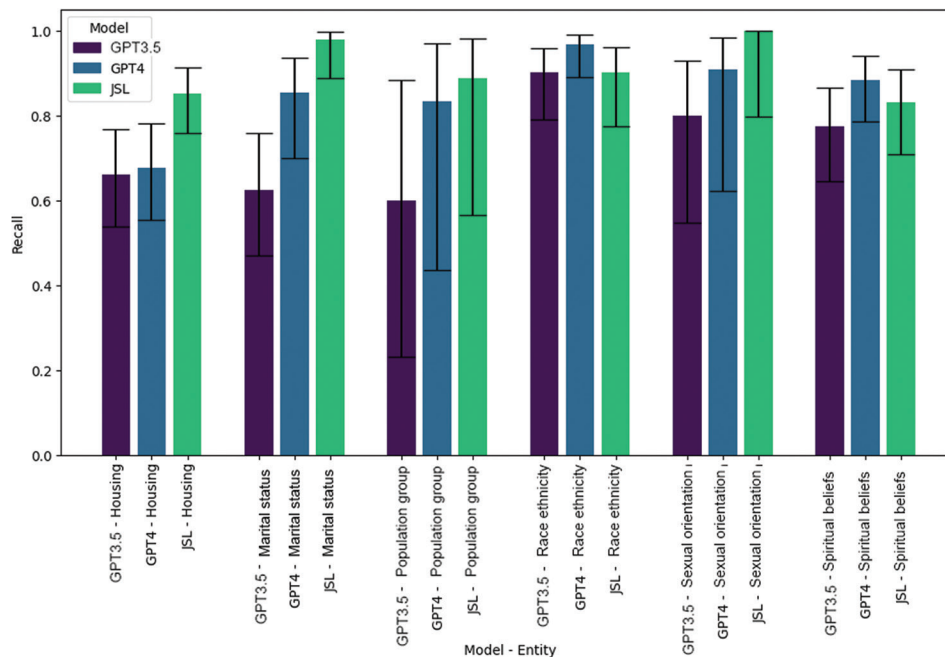


Figure S1. Comparison of recall (sensitivity; with 95% confidence intervals) between the John Snow Lab social determinants of health named-entity recognition model, generative pre-trained transformer (GPT)-3.5 (zero-shot learning), and GPT-4 (zero-shot learning) for homelessness/housing, marital status, immigrant status/population group, race/ethnicity, and spiritual belief mentions over a manually annotated corpus of clinical progress notes. The figure was plotted with Matplotlib.

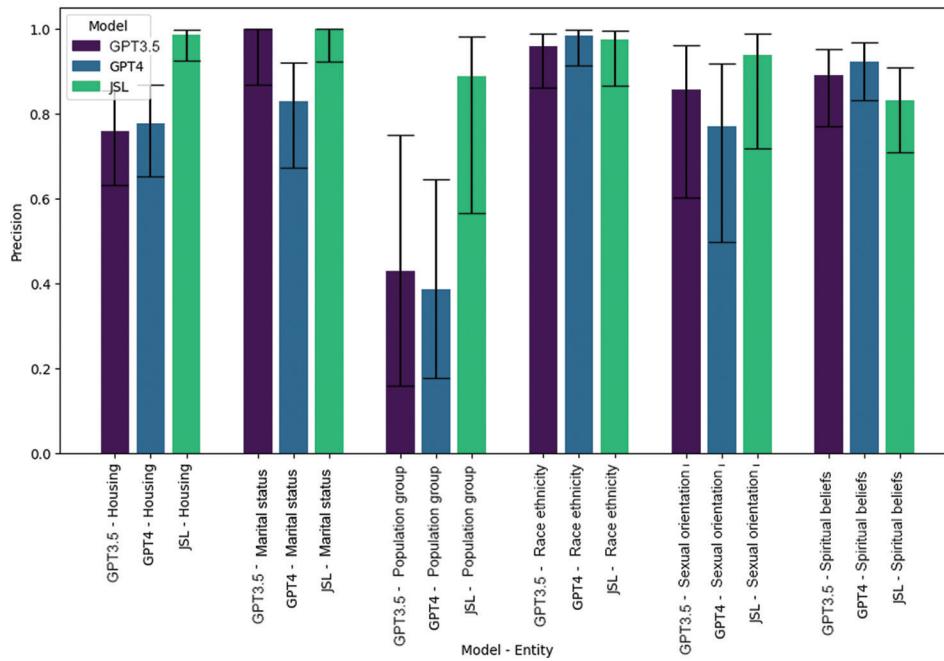


Figure S2. Comparison of precision (positive predictive value; with 95% confidence intervals) between the John Snow Lab social determinants of health named-entity recognition model, generative pre-trained transformer (GPT)-3.5 (zero-shot learning) and GPT-4 (zero-shot learning) for homelessness/housing, marital status, immigrant status/population group, race/ethnicity, and spiritual belief mentions over a manually annotated corpus of clinical progress notes. The figure was plotted with Matplotlib.

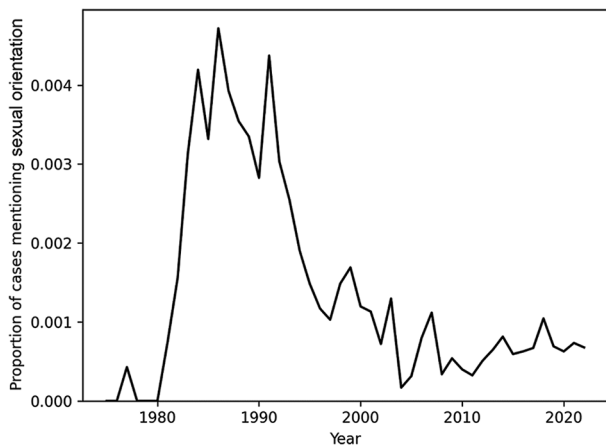


Figure S3. The proportion of clinical case reports mentioning sexual orientation from 1975 to 2022. The figure was plotted with Matplotlib.

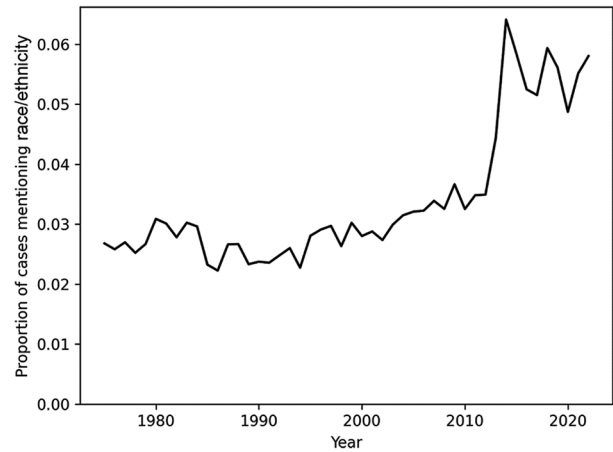


Figure S4. The proportion of clinical case reports mentioning race/ethnicity from 1975 to 2022. The figure was plotted with Matplotlib.

Table S1. Adjusted odds ratios for the probability of mentioning race/ethnicity based on publication year, clinical case type, journal specialty, journal’s geographic region, and author’s geographic region

Information	Adjusted odds ratio (95% CI)	P	Significance
Publication year	1.02 (1.02 – 1.02)	0.000	***
Clinical case type			
Hemic and lymphatic diseases	1.32 (1.26 – 1.40)	0.000	***
Eye diseases	1.30 (1.20 – 1.41)	0.000	***
Stomatognathic diseases	1.29 (1.17 – 1.43)	0.000	***
Nutritional and metabolic diseases	1.24 (1.16 – 1.32)	0.000	***
Skin and connective tissue diseases	1.21 (1.13 – 1.29)	0.000	***
Infections	1.20 (1.15 – 1.25)	0.000	***
Congenital and neonatal diseases	1.05 (0.96 – 1.13)	0.279	
Musculoskeletal diseases	1.05 (0.99 – 1.11)	0.109	
Endocrine system diseases	1.04 (0.94 – 1.15)	0.441	
Urogenital diseases	1.04 (0.98 – 1.10)	0.216	
Neoplasms	0.97 (0.94 – 1.01)	0.185	
Digestive system diseases	0.91 (0.86 – 0.97)	0.002	**
Pathological signs and symptoms	0.88 (0.85 – 0.92)	0.000	***
Nervous system diseases	0.84 (0.80 – 0.88)	0.000	***
Respiratory tract diseases	0.84 (0.78 – 0.90)	0.000	***
Cardiovascular diseases	0.77 (0.73 – 0.82)	0.000	***
Wounds and injuries	0.64 (0.58 – 0.70)	0.000	***
Occupational diseases	0.59 (0.38 – 0.91)	0.017	*
Journal specialty			
General medicine	2.18 (2.08 – 2.29)	0.000	***
Laboratory	2.10 (1.94 – 2.28)	0.000	***
Dentistry	1.97 (1.67 – 2.33)	0.000	***
Dermatology	1.82 (1.69 – 1.97)	0.000	***
Nephrology	1.53 (1.32 – 1.78)	0.000	***
Infectiology	1.51 (1.37 – 1.67)	0.000	***
Pathology	1.50 (1.39 – 1.63)	0.000	***
Rheumatology	1.48 (1.27 – 1.72)	0.000	***

(Cont’d...)

Table S1. (Continued)

Information	Adjusted odds ratio (95% CI)	P	Significance
Hematology	1.43 (1.26 – 1.62)	0.000	***
Ophthalmology	1.40 (1.27 – 1.55)	0.000	***
Psychiatry	1.39 (1.15 – 1.67)	0.000	***
Genetics	1.37 (1.01 – 1.85)	0.045	*
Gastroenterology	1.07 (0.96 – 1.20)	0.230	
Neurology	0.79 (0.72 – 0.87)	0.000	***
Pneumology	0.75 (0.61 – 0.92)	0.006	**
Cardiology	0.63 (0.56 – 0.72)	0.000	***
Surgery	0.61 (0.57 – 0.65)	0.000	***
Traumatology	0.59 (0.50 – 0.70)	0.000	***
Urology	0.48 (0.40 – 0.56)	0.000	***
Radiology	0.36 (0.30 – 0.43)	0.000	***
Rehabilitation	0.31 (0.18 – 0.54)	0.000	***
Anesthesiology	0.27 (0.20 – 0.37)	0.000	***
Journal’s geographic region			
Australia and Oceania	1.34 (1.17 – 1.53)	0.000	***
Western Europe	1.30 (1.18 – 1.43)	0.000	***
North America	0.91 (0.82 – 1.01)	0.074	
Sub-Saharan Africa	0.88 (0.74 – 1.05)	0.148	
Middle East	0.79 (0.65 – 0.96)	0.018	
North Africa	0.56 (0.35 – 0.90)	0.018	
South America	0.55 (0.43 – 0.69)	0.000	***
Eastern Europe	0.54 (0.45 – 0.64)	0.000	***
East Asia	0.48 (0.43 – 0.54)	0.000	***
Central America	0.34 (0.14 – 0.82)	0.016	*
Author’s geographic region			
Sub-Saharan Africa	4.47 (3.96 – 5.04)	0.000	***
Caribbean	3.31 (2.24 – 4.89)	0.000	***
Southeast Asia	2.89 (2.58 – 3.25)	0.000	***
East Asia	2.00 (1.90 – 2.09)	0.000	***
North America	1.77 (1.68 – 1.86)	0.000	***
Western Europe	0.92 (0.88 – 0.97)	0.002	**
Eastern Europe	0.90 (0.79 – 1.03)	0.118	
Middle East	0.77 (0.70 – 0.84)	0.000	***
Indian subcontinent	0.69 (0.62 – 0.76)	0.000	***

Notes: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.
Abbreviation: CI: Confidence interval.

Table S2. Adjusted odds ratios for the probability of mentioning marital status based on publication year, clinical case type, journal specialty, journal's geographic region, and author's geographic region

Information	Adjusted odds ratio (95% CI)	P	Significance
Publication year	1.00 (0.99 – 1.00)	0.515	
Clinical case type			
Mental disorders	2.14 (1.72 – 2.66)	0.000	***
Urogenital diseases	1.68 (1.47 – 1.91)	0.000	***
Congenital and neonatal diseases	1.08 (0.87 – 1.34)	0.513	
Pathological signs and symptoms	0.77 (0.69 – 0.86)	0.000	***
Nervous system diseases	0.71 (0.63 – 0.81)	0.000	***
Musculoskeletal diseases	0.70 (0.58 – 0.85)	0.000	***
Stomatognathic diseases	0.69 (0.48 – 0.98)	0.039	*
Nutritional and metabolic diseases	0.67 (0.52 – 0.85)	0.001	**
Respiratory tract diseases	0.67 (0.54 – 0.82)	0.000	***
Skin and connective tissue diseases	0.67 (0.53 – 0.86)	0.002	**
Hemic and lymphatic diseases	0.63 (0.51 – 0.78)	0.000	***
Infections	0.63 (0.55 – 0.73)	0.000	***
Neoplasms	0.61 (0.55 – 0.69)	0.000	***
Digestive system diseases	0.54 (0.44 – 0.67)	0.000	***
Wounds and injuries	0.51 (0.38 – 0.69)	0.000	***
Journal specialty			
Psychiatry	2.63 (1.97 – 3.51)	0.000	***
Gynecology	2.45 (2.01 – 2.99)	0.000	***
Rehabilitation	2.39 (1.31 – 4.35)	0.005	**
Forensic	2.04 (1.32 – 3.17)	0.001	**
Dermatology	1.37 (1.08 – 1.75)	0.010	*
General medicine	1.13 (1.00 – 1.27)	0.046	*
Surgery	0.74 (0.64 – 0.86)	0.000	***
Gastroenterology	0.53 (0.33 – 0.84)	0.007	**
Traumatology	0.46 (0.26 – 0.79)	0.005	**
Rheumatology	0.37 (0.15 – 0.89)	0.027	*
Nephrology	0.21 (0.09 – 0.51)	0.001	**
Journal's geographic region			
Indian subcontinent	1.61 (1.22 – 2.13)	0.001	**
East Asia	0.85 (0.73 – 0.98)	0.029	*
Eastern Europe	0.65 (0.46 – 0.91)	0.011	*
Author's geographic region			
Sub-Saharan Africa	1.98 (1.32 – 2.96)	0.001	**
Southeast Asia	1.75 (1.19 – 2.58)	0.005	**
North Africa	1.73 (1.06 – 2.80)	0.027	*
Indian subcontinent	1.49 (1.19 – 1.87)	0.000	***
Middle East	1.31 (1.09 – 1.58)	0.004	**
East Asia	0.89 (0.78 – 1.02)	0.100	
Western Europe	0.88 (0.78 – 1.00)	0.047	*
Central America	0.30 (0.07 – 1.21)	0.090	

Notes: *P<0.05; **P<0.01; ***P<0.001.
Abbreviation: CI: Confidence interval.

Table S3. Adjusted odds ratios for the probability of mentioning sexual orientation based on publication year, clinical case type, journal specialty, journal's geographic region, and author's geographic region

Information	Adjusted odds ratio (95% CI)	P	Significance
Publication year	0.96 (0.95 – 0.96)	0.000	***
Clinical case type			
Infections	25.00 (19.68 – 31.75)	0.000	***
Mental disorders	1.73 (0.91 – 3.31)	0.097	
Urogenital diseases	0.69 (0.46 – 1.02)	0.063	
Pathological signs and symptoms	0.66 (0.53 – 0.84)	0.001	**
Nervous system diseases	0.57 (0.41 – 0.78)	0.000	***
Endocrine system diseases	0.47 (0.15 – 1.46)	0.189	
Respiratory tract diseases	0.37 (0.21 – 0.65)	0.000	***
Musculoskeletal diseases	0.30 (0.15 – 0.58)	0.000	***
Nutritional and metabolic diseases	0.30 (0.11 – 0.80)	0.016	*
Wounds and injuries	0.29 (0.09 – 0.90)	0.032	*
Congenital and neonatal diseases	0.28 (0.09 – 0.88)	0.030	*
Cardiovascular diseases	0.26 (0.14 – 0.47)	0.000	***
Otorhinolaryngologic diseases	0.22 (0.03 – 1.56)	0.130	
Immune system diseases	0.21 (0.05 – 0.83)	0.026	*
Journal specialty			
Psychiatry	2.27 (1.02 – 5.04)	0.044	
Dermatology	1.65 (1.21 – 2.25)	0.002	**
Neurology	1.65 (1.14 – 2.38)	0.008	**
Pathology	1.60 (1.11 – 2.31)	0.012	*
General medicine	1.25 (0.98 – 1.59)	0.073	
Infectiology	1.22 (0.91 – 1.65)	0.187	
Surgery	0.46 (0.30 – 0.69)	0.000	***
Radiology	0.19 (0.05 – 0.77)	0.020	*
Pediatrics	0.16 (0.07 – 0.39)	0.000	***
Traumatology	0.15 (0.02 – 1.08)	0.060	
Cardiology	0.13 (0.02 – 0.92)	0.041	
Journal's geographic region			
South America	4.04 (1.92 – 8.50)	0.000	***
North America	2.15 (1.31 – 3.55)	0.003	**
Western Europe	1.64 (1.00 – 2.68)	0.049	*
Eastern Europe	1.01 (0.40 – 2.56)	0.991	
East Asia	0.95 (0.53 – 1.71)	0.867	
Author's geographic region			
North America	1.47 (1.13 – 1.91)	0.004	**
Western Europe	1.46 (1.15 – 1.87)	0.002	**
East Asia	0.63 (0.44 – 0.91)	0.013	*
Indian subcontinent	0.16 (0.04 – 0.63)	0.009	**
Middle East	0.16 (0.05 – 0.51)	0.002	**

Notes: *P<0.05; **P<0.01; ***P<0.001.
Abbreviation: CI: Confidence interval.

Table S4. Adjusted odds ratios for the probability of mentioning immigrant status based on publication year, clinical case type, journal specialty, journal’s geographic region, and author’s geographic region

Information	Adjusted odds ratio (95% CI)	p	Significance
Publication year	1.00 (0.99 – 1.01)	0.706	
Clinical case type			
Infections	4.51 (3.53 – 5.77)	0.000	***
Mental disorders	2.05 (1.14 – 3.71)	0.017	*
Pathological signs and symptoms	0.68 (0.51 – 0.91)	0.009	**
Skin and connective tissue diseases	0.68 (0.34 – 1.34)	0.261	
Eye diseases	0.66 (0.26 – 1.64)	0.371	
Nervous system diseases	0.64 (0.44 – 0.94)	0.021	*
Neoplasms	0.44 (0.31 – 0.63)	0.000	***
Cardiovascular diseases	0.43 (0.27 – 0.69)	0.000	***
Immune system diseases	0.20 (0.03 – 1.46)	0.114	
Journal specialty			
Gynecology	4.25 (2.64 – 6.82)	0.000	***
Psychiatry	3.94 (1.95 – 7.95)	0.000	***
General medicine	1.75 (1.34 – 2.29)	0.000	***
Neurology	0.64 (0.31 – 1.34)	0.239	
Dermatology	0.28 (0.09 – 0.90)	0.033	*
Ophthalmology	0.12 (0.02 – 0.92)	0.042	*
Journal’s geographic region			
Middle East	2.20 (1.19 – 4.07)	0.012	*
East Asia	0.32 (0.15 – 0.65)	0.002	**
Author’s geographic region			
Australia and Oceania	2.17 (1.14 – 4.12)	0.018	*
North America	1.53 (1.17 – 2.01)	0.002	**
East Asia	0.23 (0.12 – 0.45)	0.000	***
South America	0.14 (0.02 – 1.02)	0.053	
Indian subcontinent	0.09 (0.01 – 0.62)	0.015	*

Notes: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.
Abbreviation: CI: Confidence interval.

Table S5. Adjusted odds ratios for the probability of mentioning homelessness based on publication year, clinical case type, journal specialty, and journal’s geographic region

Information	Adjusted odds ratio (95% CI)	p	Significance
Publication year	1.05 (1.02 – 1.08)	0.000	***
Clinical case type			
Infections	6.36 (3.57 – 11.32)	0.000	***
Mental disorders	5.80 (2.26 – 14.89)	0.000	***
Wounds and injuries	4.73 (2.29 – 9.77)	0.000	***
Chemically-induced disorders	2.70 (0.94 – 7.73)	0.064	
Neoplasms	0.52 (0.22 – 1.19)	0.119	
Journal specialty			
Forensic	14.92 (5.48 – 40.64)	0.000	***
Pathology	3.95 (1.39 – 11.28)	0.010	*
Infectiology	3.75 (1.77 – 7.94)	0.001	**
Journal’s geographic region			
Eastern Europe	4.76 (1.88 – 12.03)	0.001	**

Notes: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.
Abbreviation: CI: Confidence interval.

Table S6. Adjusted odds ratios for the probability of mentioning the patients’ spiritual beliefs based on publication year, clinical case type, journal specialty, journal’s geographic region, and author’s geographic region

Information	Adjusted odds ratio (95% CI)	P	Significance
Publication year	1.01 (0.99 – 1.03)	0.498	
Clinical case type			
Endocrine system diseases	3.47 (1.38 – 8.68)	0.008	**
Mental disorders	3.05 (1.27 – 7.31)	0.012	*
Cardiovascular diseases	0.42 (0.15 – 1.16)	0.094	
Respiratory tract diseases	0.23 (0.03 – 1.69)	0.150	
Neoplasms	0.20 (0.08 – 0.50)	0.001	**
Journal specialty			
Psychiatry	7.61 (2.93 – 19.79)	0.000	***
Surgery	0.23 (0.06 – 0.96)	0.044	*
Journal’s geographic region			
Middle East	5.05 (1.99 – 12.85)	0.001	**
Author’s geographic region			
Sub-Saharan Africa	9.17 (2.84 – 29.64)	0.000	***
Indian subcontinent	4.09 (1.83 – 9.15)	0.001	**

Notes: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.
Abbreviation: CI: Confidence interval.