



# Using keyword text analysis and unsupervised machine learning to manage large search results

## What

### A scoping review

- Community-based research
- Underserved populations
- Engagement

## The Problem

### Complex search

- Underserved populations difficult to define; context diffuse across disciplines

disabilities  
disability  
disabled  
disadvantage  
disadvantaged  
disparities

**diverse**

diversity  
elderly  
equitable  
equities  
equity

**ethnic**

**gender**

HIV/AIDS  
homeless  
homelessness  
immigrant  
immigrants  
indigent

**inequality**

inequities  
inequity  
lesbian  
LGBT  
LGBTQ  
low income  
marginal  
marginalization  
marginalized

medically complex  
mental illness  
mentally disabled  
mentally ill  
migrant

**minorities**

minority  
oppressed  
oppression  
peripheralized  
pregnant women  
refugee

**rural**

sexual orientation  
sexuality

**socioeconomic**

stigma  
stigmatized  
stigmatizing  
transgender  
transgendered  
underinsured  
underpopulated  
underrepresented  
under-represented  
underserved  
uninsured  
veteran  
veterans

**vulnerable**

Keyword	Percentage Occurrence	Number of Documents with Keyword	Number of Unique Documents with Keyword
gender	13.62	3503	1569
rural	12.99	3339	1748
diverse	12.58	3234	1705
socioeconomic	5.71	1467	500
vulnerable	5.6	1440	534
stigma	5.47	1407	421
ethnic	5.36	1377	205
low income	5.32	1367	481
elderly	5.31	1365	846
minority	5.21	1340	255
ethnicity	3.73	958	231
disabilities	3.16	813	314
mental illness	3.1	797	357
pregnant women	3.08	793	451
discrimination	3.04	782	195
equity	2.47	634	194
underserved	2.41	620	143
vulnerability	2.31	595	241
gay	2.18	560	69
inequalities	1.87	480	116
disadvantaged	1.86	479	109
bisexual	1.85	475	11
veterans	1.82	468	229

**% occurrence:**  
percent of total results

**# documents with keyword:**  
raw count out of total results

**# unique docs w/ keyword:**  
times document appears *only*  
because of this keyword in the set  
(co-occurrence)

## Research aim: engaging vulnerable populations in community-based research

Cluster	# of studies	Topic Signature / Keywords
1	4334	<b>Cluster 1:</b> research   public   policy   social   development   review   local   countries   issues   researchers   process   based   paper   ethical   participation   communities   new   equity   evidence   studies
2	2011	
3	2656	
4	1319	ci   95   odds   associated   risk   ratio   confidence   years   interval   age   participants   adjusted   aor   women   prevalence   factors   regression   higher   association   compared
5	997	cancer   screening   breast   cervical   women   american   research   disparities   patients   care   colorectal   african   based   participants   education   intervention   survivors   underserved   patient   knowledge
6	2797	mental   qualitative   stigma   participants   experiences   women   interviews   social   focus   care   people   support   barriers   themes   illness   research   groups   services   family   findings
7	4181	<b>Cluster 8:</b> research   students   participatory   based   cbpr   program   academic   disparities   communities   partnership   partnerships   project   learning   development   university   training   education   researchers   process   approach
8	1894	
9	2136	
10	1358	intervention   trial   randomized   controlled   group   control   based   participants   months   program   interventions   baseline   outcomes   effectiveness   physical   children   care   activity   design   follow