Phenome-Wide Associations of Homelessness in Electronic Health Records

Jared M. Hendrickson BS[1], Joshua C. Denny, MD, MS[2], Cosmin A. Bejan, PhD[2]
[1]Northeastern University, Boston, MA; [2]Vanderbilt University Medical Center, Nashville, TN

Introduction Few studies that leverage the information available in Electronic Health Records (EHRs) have been conducted to assess the impact of socio-behavioral determinants of health on the overall quality of life. Homelessness is a major environmental stress on an individual and is known to affect many diseases and access to care. We investigated associations between homelessness and a wide range of clinical phenotypes using a large-scale EHR repository.

Methods For this study, we identified patients with eras of homelessness from the Vanderbilt EHR using text mining and administrative billing codes (ICD9:V60.0, ICD10:Z59.0).[1] The text mining algorithm had a positive predictive value of 93%. We 1:1 matched the homeless individuals to patients without homelessness status by age, sex, and race. To investigate homelessness associations over the entire phenome, we used logistic regression models adjusted for age, sex, and race in a PheWAS framework.[2] Additionally, we performed stratified analyses by race, sex, and disease onset with respect to homelessness. Specifically, for the temporal analysis of a disease of interest, we stratified the homeless patients into the before/after stratum based on the patient's first episode of homelessness in the EHR. To better account for small sample sizes, the Fisher's exact test was applied to each phenotype for each stratification combination of race, sex, and temporal homelessness status.

Results We extracted a cohort of 3165 homeless patients, which includes 1591 European-American (EA) males, 654 EA females, 626 African American (AA) males, and 238 AA females. Preliminary data analysis revealed that the homeless group exhibited on average four times as many ICD codes per patient as the matched never homeless cohort ($p=4.5\times10^{-118}$). The top ranked diseases by odds ratios (OR) in the adjusted logistic regression models include suicidal ideation (OR=145, 95% CI=[74.7, 281.2], $p=5.9\times10^{-49}$), alcoholism (OR=91, 95% CI=[64.3, 129.3], $p=8.1\times10^{-142}$), personality disorders (OR=80, 95% CI=[29.3, 216.1], $p=8.4\times10^{-18}$), schizophrenia (OR=78, 95% CI=[43.7, 139.5], $p=4.1\times10^{-49}$), and substance abuse (OR=67, 95% CI=[51.8, 87.8], $p=1.7\times10^{-214}$). Among all race and sex combination strata in the before homeless temporal cohort, suicidal ideation and attempt, substance addiction and disorders, mood disorders, depression, and anxiety were all found significant after Bonferroni correction and a minimum case threshold of 40. Notable associations from our study include a high odds ratio for HIV among AA males before homelessness episodes (OR=11.4, 95% CI=[5.1, 29.8], $p=2.0\times10^{-14}$) versus EA males (OR=2.7, 95% CI=[1.5, 5.2], $p=7.2\times10^{-4}$). Using strata analysis and the Breslow-Day test, homeless status (p=0.003), race (p=0.003) but not sex (p=0.11) were found to be significant variables of interaction for HIV status in AA males. While most significant phenotypes continued from the before homeless to the after homeless temporal cohort stratification, few novel and significant associations stemming from homelessness status were found. Via a phenome-wide Fisher's exact test for EA males, the largest race-sex combination cohort, alcoholic liver damage (OR=8.2, 95% CI=[3.6, 19.8], $p=4.1\times10^{-8}$), acid-base balance disorder (OR=2.5, 95% CI=[1.2, 5.9], $p=1.6\times10^{-2}$), chronic pain (OR=2.7, 95% CI=[1.5, 5.3], $p=5.4\times10^{-4}$), and arrhythmias (OR=2.8, 95% CI=[1.3, 6.2], $p=5.4\times10^{-3}$) were discovered significant in the after but not the before homeless status patients.

Conclusion From a comparative analysis of previous findings, significant associations in the homeless population were replicated for alcoholism, depression, HIV status, tuberculosis, and schizophrenia.[3] Our data supports the trend of destabilization, especially in mental health, with impending homelessness. It is important to note that the time at the first homeless status recorded in EHR may be different from the timestamp of the first homeless episode and that the subsequent homeless status cohort represents a generally healthier cohort than the impending homeless patient cohort due to the influx of homeless individuals by social factors rather than medical reasons. For future studies, larger sample sizes would help clarify associations, especially in minority groups and low prevalence phenotypes.

References

1. Bejan CA, *et al*. Mining 100 million notes to find homelessness and adverse childhood experiences: 2 case studies of rare and severe social determinants of health in electronic health records. J Am Med Inform Assoc 2018; 25:61–71.
2. Denny JC, *et al*. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 2010; 26:1205–10.
3. Wagner J, *et al*. Health status and utilisation of the healthcare system by homeless and non-homeless people in Vienna. Health Soc Care Community 2014; 22:300–7.