



Published in final edited form as:

J Am Coll Cardiol. 2020 December 15; 76(24): 2862–2874. doi:10.1016/j.jacc.2020.10.023.

Unconventional Natural Gas Development and Hospitalization for Heart Failure in Pennsylvania

Tara P. McAlexander, PhD, MPH^{1,2}, Karen Bandeen-Roche, PhD, MS^{3,4}, Jessie P. Buckley, PhD, MPH^{1,5}, Jonathan Pollak, MS¹, Erin D. Michos, MD, MHS⁴, John William McEvoy, MBBCh, MHS^{4,6}, Brian S. Schwartz, MD, MS^{1,4,5}

¹Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States ²Department of Epidemiology and Biostatistics, Drexel University Dornsife School of Public Health, Philadelphia, Pennsylvania, United States ³Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States ⁴Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States ⁵Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States ⁶National Institute for Preventive Cardiology, National University of Ireland, Galway, Ireland

Abstract

Background: Growing literature linking unconventional natural gas development (UNGD) to adverse health has implicated air pollution and stress pathways. Persons with heart failure (HF) are susceptible to these stressors.

Objectives: To evaluate associations between UNGD activity and hospitalization among HF patients; stratified both by ejection fraction status (reduced [HF_rEF], preserved [HF_pEF], not classifiable) and by HF severity.

Methods: We evaluated the odds of hospitalization among patients with HF seen at Geisinger from 2008–2015 using electronic health records. We assigned metrics of UNGD activity by phase (pad preparation, drilling, stimulation, and production) 30 days prior to hospitalization or frequency-matched control selection date; we assigned phenotype status using a validated algorithm.

Results: We identified 9,054 HF patients with 5,839 hospitalizations, with a mean (SD) age of 71.1 (12.7) years; 47.7% were female. Comparing the 4th to 1st quartiles, adjusted odds ratios (95% CI) for hospitalization were 1.70 (1.35 – 2.13), 0.97 (0.75–1.27), 1.80 (1.35–2.40), and 1.62 (1.07–2.45) for the pad preparation, drilling, stimulation, and production metrics, respectively. We

Address for correspondence: Brian S. Schwartz, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Room W7041, Baltimore, Maryland 21205, Telephone: 410-955-4158 (phone), Fax: 410-955-1811 (fax), bschwar1@jhu.edu, Twitter: @TaraMc137.

Competing financial interest disclosure: The authors declare that they have no competing financial interests.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

did not find effect modification by HF \neq EF or HF p EF status. Associations of most UNGD metrics with hospitalization were stronger among those with more severe HF at baseline.

Conclusions: Three of four phases of UNGD activity were associated with hospitalization for HF in a large sample of HF patients in an area of active UNGD, with similar findings by HF \neq EF vs HF p EF status. Older patients with HF appear particularly vulnerable to adverse health impacts from UNGD activity.

Condensed abstract:

This nested case-control study evaluated the adjusted odds of hospitalization among patients with heart failure (HF) in relation to four phases of unconventional natural gas development (UNGD) activity in the month prior. This study also evaluated whether heart failure phenotype (HF \neq EF vs. HF p EF) and a measure of heart failure severity modified the associations between UNGD activity and hospitalization for HF. We found that HF patients in the highest compared to lowest quartiles of UNGD activity had an increased likelihood of hospitalization for HF. This association was stronger for more severe HF patients, suggesting that UNGD activity could exacerbate HF.

Keywords

Heart failure; Hospitalization; Unconventional natural gas development; Environment

Introduction

Heart failure is a common chronic condition affecting over 25 million persons globally and 5.7 million Americans (1–4), costing the United States (US) health care system over \$30 billion annually (2,5,6). There are two main heart failure phenotypes: reduced (HF \neq EF) or preserved ejection fraction (HF p EF) (7,8). Both lead to impaired cardiac output, reducing blood flow to critical organs (9). This makes patients with heart failure at risk for hospitalizations and mortality (10) and susceptible to environmental exposures (11). Increasingly, epidemiologic studies report associations of air pollutants (12–15) and noise (13) with hospital admissions for heart failure, thought to be due to systemic inflammation, direct tissue injury, ischemia, arrhythmias, or thrombosis (11,16).

Unconventional natural gas development (UNGD) is a growing industry worldwide, and Pennsylvania in the US has seen over 12,000 wells drilled in the Marcellus shale since 2004 (17,18). UNGD has a number of environmental impacts, including noise (19) and air pollution (e.g., PM_{2.5}, oxides of nitrogen [NO_x], oxides of sulfur [SO_x], volatile organic compounds [VOCs], and polycyclic aromatic hydrocarbons [PAHs]) associated with its several stages (20–23): preparation of well pads, drilling, stimulation (i.e., hydraulic fracturing, referred to as “fracking”) and production. Combined and cumulative impacts of UNGD can adversely affect psychosocial stress and community well-being (24–26).

The purpose of this study was to evaluate associations between UNGD activity metrics, by phase of development, and hospitalization among patients with heart failure, overall and by HF p EF and HF \neq EF phenotypes. In *post hoc* analysis, a measure of baseline heart failure severity was used to evaluate whether associations between UNGD activity metrics and

hospitalization differed by severity. To our knowledge, no prior epidemiologic studies have examined associations between measures of UNGD activity and heart failure outcomes.

Methods

Study population and design

We conducted a case-control study among heart failure patients, comparing persons with and without hospitalizations, using electronic health record (EHR) data from Geisinger, an integrated health system with multiple inpatient and outpatient centers in Pennsylvania, for January 1, 2008 to July 31, 2015, coinciding with active UNGD. The study was approved by Geisinger's institutional review board (IRB) and nested within the general-population-representative, open, dynamic cohort among persons that have a Geisinger primary care provider (27). We identified heart failure diagnoses codes from inpatient, outpatient, or emergency department (ED) encounter records and medication order records, excluding laboratory orders and the problem list. We identified 16,098 patients with a heart failure diagnosis code (International Classification of Diseases [ICD]-9 428.x). After exclusions (Figure 1), there were 12,330 individuals eligible for selection.

Case identification

We identified 5,839 of these 12,330 patients who were hospitalized for heart failure. We included only incident heart failure hospitalizations.

Control selection

Patients with heart failure were eligible for control selection if they did not have a hospitalization for heart failure at Geisinger prior to the date of selection as a control. The rationale was to evaluate a control group as similar as possible to the cases except for the heart failure hospitalization event, as our goal was to evaluate if UNGD activity was associated with an exacerbation. To limit the potential for confounding by age, sex, and year, we used 1:1 incidence density sampling with replacement of selected controls and frequency-matched cases to control encounters by these variables (28). Control encounters included any outpatient visit or medication order that did not include a heart failure diagnosis code for that date. Controls were selected once per year in the year of the case's hospitalization for a maximum of five control encounters over the study period. There were 9,054 persons in the analysis; 3,215 only served as controls.

Covariate assignment

Information for time-invariant patient characteristics (e.g., sex, race/ethnicity) were available from the EHR. We calculated time-varying covariates at hospitalization or control selection date: age, smoking status (never, previous, current), Charlson index of morbidity (29), and receipt of Medical Assistance (a surrogate for family socioeconomic status (30,31)). We identified co-morbid conditions based on at least two encounter diagnosis codes on any date between January 1, 2008 and the hospitalization or control encounter date.

We calculated duration of care as duration from patients' first contact with Geisinger and date of hospitalization or control date. We identified current medication use by verifying that

the date of hospitalization or control encounter was between the start and end dates of the medication order (32–34).

We used available height and weight measurements to calculate body mass index (BMI, kg/m²) at the date closest to either the heart failure hospitalization or control selection date. For individuals without sufficient height and weight data to calculate BMI within one year of the case event or control encounter date (n = 339, 3.7% of the 9,054 study patients), we imputed BMI for these patients using multiple imputation based on age, sex, and receipt of Medical Assistance (35).

Heart failure phenotype assignment

Heart failure phenotypes from the Electronic Medical Records and Genomics (eMERGE) Network had been previously assigned for patients who had notation of heart failure in the EHR problem list or encounter notes, a diagnosis for heart failure, and measurement of ejection fraction (36,37). We evaluated four phenotypic groups: HF_pEF, HF_rEF, those without enough information to be phenotyped (“eMERGE not applied”), and those who had information to have the eMERGE algorithm applied but did not have a clear HF_pEF or HF_rEF phenotype (“eMERGE no phenotype”). The primary analysis using phenotype evaluated effect modification of the relations of UNGD metrics with hospitalization by HF_pEF vs. HF_rEF status.

In early analysis of the phenotype data, we noticed that patients identified as HF_pEF or HF_rEF by the eMERGE algorithm (vs. those who were not): 1) had a greater proportion of patients deceased by the end of the study period; 2) were more likely to be hospitalized during the study period; 3) had a higher proportion of persons taking antihypertensive, antihyperlipidemic, and anticoagulant medications; 4) had a higher proportion of comorbid diagnoses relevant to heart failure; and 5) had a higher mean Charlson index. We interpreted this as evidence that phenotyped patients had more severe heart failure than non-phenotyped patients. Thus, in a *post hoc* analysis, we dichotomized the eMERGE phenotypic groups into phenotyped vs. not as a measure of disease severity and evaluated effect modification of the primary associations by this measure.

Assignment of community metrics

We used patients’ residential addresses to obtain latitude and longitude coordinates as previously reported (30,38,39) and located the residence in communities as townships, boroughs, or census tracts in cities (40–42). Residential locations were grouped into five sub-regions of our 38-county study area consisting of 5 to 15 contiguous counties each (northeast, southeast, central, southwest, and northwest) to account for potential spatial confounding at a larger scale. We calculated a measure of community socioeconomic deprivation (CSD) using six indicators from 2010–2014 data from the US Census American Community Survey (39–43).

We obtained the locations of major (i.e., highways) and minor (i.e., arterial and local) roads from the Federal Highway Administration and calculated the Euclidian distance (meters) from each patient’s residential address to the nearest road. We calculated the distance (m) from patients’ residential addresses to the nearest Geisinger hospital or clinic to assess

proximity to care. To evaluate associations of greenness with hospitalization, we obtained NASA MODIS satellite data for 16-day periods of maximum greenness for each year from 2008–2015. We assigned the normalized difference vegetation index (NDVI), a measure of greenness, to subjects based on the NDVI values in the 1250 m x 1250 m grid surrounding their residential address.

UNGD activity assignment

UNGD activity metric data and calculation have been previously reported (39). We used data from the Pennsylvania Department of Environmental Protection on UNGD wells for 2008 to 2015, documenting dates and locations of four phases of activity: well pad preparation (e.g., clearing of site, delivery of equipment and personnel), well drilling (i.e., starting at the spud date), well stimulation (hydraulic fracturing), and natural gas production. We assigned activity metrics that incorporated number, phase, size, and location of wells, and divided by the squared distance from residential locations to all wells in the state (38,39,44–46). UNGD activity assignments used the following equation (46) where j identified patient, n was the number of wells, and d^2_{ij} was the squared distance (m^2) between patient j 's residential address and well i :

$$\text{Activity metric for patient } j = \sum_{i=1}^n \frac{m_i}{d^2_{ij}}$$

For the pad preparation and spud activity metrics, $m_i = 1$. For the stimulation and production metrics, m_i was total well depth (m) or total daily volume of natural gas (m^3) produced for well i , respectively (46). We calculated each of these activity metrics for a duration of 30 days with a one-day lag prior to hospitalization or control encounter date (Supplemental Figure 1), hypothesizing that exposures, which encompassed more than just air pollution, required 30 days' duration to contribute to hospitalization (24,47).

Statistical methods

There were three primary goals of the analysis: first, to separately evaluate associations between four metrics of UNGD activity and hospitalization for heart failure; second, to evaluate effect modification of these associations by HFpEF vs. HF \neq EF status; and third, in a *post hoc* analysis, to evaluate effect modification of the association between metrics of UNGD activity and hospitalization for heart failure by heart failure severity. All analyses were conducted using Stata v13.1 (StataCorp LP 2016. Stata/MP 13.1), R (R Core Team, 2014), or ArcGIS (ESRI 2011. ArcGIS Desktop: Release 10.3).

We compared the distribution of individual-level covariates, community metrics, and UNGD activity metrics for cases and controls. We used the *melogit* function in Stata v13.1 to develop multilevel logistic regression models estimating the odds of hospitalization, comparing cases to controls, by quartile of UNGD activity. We included random intercepts for patient (to account for correlation within individuals over time who were included in analysis more than once as control then as case) and community (to account for the correlation of measures for persons clustered in communities). We evaluated non-linearity for continuous variables by evaluating linear, quadratic, and cubic terms after centering the

variable; higher order terms were only included if the association crossed an inferential boundary ($p < 0.05$).

Our initial model included *a priori* variables sex (female vs. male), age category, race/ethnicity (nonwhite vs. white), BMI, and Medical Assistance based on evidence of risk of heart failure hospitalization (48). Additional models evaluated, in a stepwise fashion, year of hospitalization or control selection date; geographic region; year and region; season, duration of health care contact, distance to nearest hospital or clinic; and distance to both major and minor roads. Aside from the *a priori* variables, we retained variables in the model if they changed the effect estimates for any of the four UNGD activity metrics by more than 5%.

Evaluation of effect modification by heart failure phenotype (HFpEF vs. HFrEF).

We examined adjusted multilevel logistic regression models of hospitalization that included the main effects of each phenotype indicator, quartiles of UNGD activity, and cross-product terms between phenotype categories and each UNGD metric. We calculated the global p-value of the cross-product terms for the HFpEF indicator and UNGD metrics with χ^2 tests. We generated linear combinations of both main effects and their cross-products to estimate stratum-specific odds ratios for hospitalization.

Evaluation of effect modification by heart failure severity.

We evaluated cross-products between each UNGD activity metric and the severity indicator. We calculated the global p-value of the cross-product terms with χ^2 tests and then estimated stratum-specific odds ratios for hospitalization.

Sensitivity analyses

We conducted several sensitivity analyses. First, we evaluated associations between UNGD activity and heart failure hospitalization only among patients between the ages of 40 and 80 years. Second, we evaluated the number of days since a patient's heart failure diagnosis instead of duration of contact to differently measure a patient's duration of care. Third, to evaluate whether spatial confounding could account for our observed associations, we conducted a negative exposure control analysis, assigning UNGD activity metrics in a temporally nonsensical way, such that the UNGD activity metrics could not have caused heart failure hospitalization (49,50). We limited our analyses to events from 2008 and 2009, and we assigned UNGD activity to these events from six years after the hospitalization or control selection date. Lastly, we evaluated associations with adjustment for NDVI, and with UNGD metrics modeled continuously with linear and quadratic terms.

Results

Description of study patients

Of the 12,330 patients eligible for selection into this study (Figure 1), 5,839 had incident hospitalization for heart failure between 2008–2015 and were identified as cases. In unadjusted analysis, there were small differences between case and control patients for duration of contact with the health system, distance to major and minor roads, and

distributions of community type, smoking status, and proportion of patients deceased at the end of the study period (Table 1). Comparing patients by phenotype groups, there were differences by sex, age, comorbid diagnoses, medication use, deceased status, and duration of heart failure in unadjusted analyses (Table 2).

Adjusted associations of UNGD activity metrics with heart failure hospitalization

We observed exposure-effect relations, with increasing levels of covariate control, for three of four UNGD activity metrics with the adjusted odds of heart failure hospitalization (Table 3). After adjustment for *a priori* covariates (**Model 1**), there were apparent exposure-effect relations for the stimulation and production metrics. Addition of year (**Model 2**) strengthened exposure-effect relations for pad preparation, stimulation, and production metrics. Addition of a regional indicator variable attenuated all associations (**Model 3**). When region, year (**Model 4**), observation time, distance to nearest Geisinger hospital or clinic, and season (**Model 5**), were added to models, associations with spud were not present, while pad preparation, stimulation, and production metrics evidenced exposure-effect relations. For example, the OR [95% CI] for quartiles 2–4 of the pad preparation metric were: 1.19 (1.01, 1.40), 1.63 (1.35, 1.97), and 1.70 (1.35, 2.13) compared to the first quartile.

Effect modification of associations of UNGD activity metrics with heart failure hospitalization by heart failure phenotype

We observed exposure-effect associations with increasing quartile of UNGD activity, particularly for the stimulation metric, in both heart failure phenotype groups (Table 4), with 4th quartile associations in both phenotypes for the pad preparation, stimulation, and production metrics compared to the first quartile of HF_pEF patients. Pad preparation showed nonsignificant evidence of effect modification by phenotype (p-value for cross-products = 0.05).

There was evidence that heart failure severity (i.e., eMERGE phenotyped vs. not phenotyped) modified relations of the pad and spud metrics with hospitalization (Figure 2), with respective p-values of 0.009 and 0.03. These associations were stronger in the phenotyped group for pad preparation and spud metrics.

Sensitivity analyses

Associations from several sensitivity analyses of models with: the random intercept for person removed; age restrictions; and adjustment for duration of heart failure instead of duration of contact with the health care system did not inferentially change results. The negative exposure control analysis did not reveal associations between UNGD activity metrics from 2014 and 2015 assigned to events six years before, with the exception of the 4th quartile of UNGD production activity (OR [95% CI]: 1.35 [1.00, 1.82]). Compared to the primary analyses, the negative control analysis resulted in reductions in effect size for UNGD metrics by more than 10 %. Modeling the UNGD metrics continuously (Supplemental Table 1, Supplemental Figure 2) did not inferentially change our results from the primary analysis, and only the pad preparation and production metrics evidenced non-linearity. Additional sensitivity analyses included additional adjustment for NDVI

(Supplemental Table 2) and modeling UNGD metrics as a combined z-score sum (Supplemental Table 3), and these also did not change inference of primary findings.

Discussion

Our findings suggested that individuals living with heart failure, when exposed to greater UNGD activity, are more likely to be hospitalized. These associations were stronger in individuals with more severe heart failure at baseline. This study was motivated by strong biologic rationale and *a priori* hypotheses regarding how the environmental impacts of UNGD (44,51,52) could affect cardiovascular health in older adults. This study's important findings are biologically relevant, consistent with *a priori* knowledge, and thus contribute to growing epidemiologic evidence that environmental factors can exacerbate heart failure.

We observed exposure-effect relations across quartiles of UNGD activity and odds of hospitalization for heart failure for three of the four UNGD phases evaluated. Stronger associations between the pad preparation, stimulation, and production activity are consistent with known environmental exposures associated with these phases, including increases in air pollution, traffic, and noise (20). These associations were robust to increasing spatiotemporal covariate adjustment and multiple sensitivity analyses; effect estimates of this size are unlikely to be explained by unmeasured confounding. We observed that persons with more severe heart failure had greater odds of hospitalization. We hypothesized that the HF p EF phenotype (vs. HF r EF) would modify the association between UNGD activity and hospitalization but did not find evidence of this. We did observe stronger associations with UNGD activity in these groups compared to the overall sample, but the magnitudes of these effect estimates did not appreciably differ between HF p EF and HF r EF patients, suggesting that both HF p EF and HF r EF patients are susceptible to exposures related to UNGD activity.

In evaluation of effect modification by heart failure severity, exposure-effect associations were present for all four UNGD activity metrics, with three crossing an inferential boundary. We believe this heart failure severity measure was valid, because persons who could be phenotyped (vs. not) were more likely to die, be hospitalized for heart failure, had other relevant diagnoses (e.g., myocardial infarction), were taking more medications, and had a higher Charlson index. This suggests that vulnerable persons with severe heart failure might be more susceptible to the adverse effects of UNGD activity that could lead to hospitalization. This is an important finding for several reasons: first, it underscores the importance of our primary associations between UNGD activity and hospitalization; second, it alleviates concerns that our primary results could be an artifact of spatial or temporal confounding; and third, it suggests that biologic mechanisms due to disease severity likely mediate the associations observed between increasing UNGD activity and increasing likelihood of hospitalization for heart failure.

This study utilized EHR data from a large, representative population in Pennsylvania living with varying intensity of UNGD activity over an eight-year study period. We illustrate that the adjustment of year in our primary models was important because our activity metrics varied by year. Incidence density sampling of cases and controls by year helped to alleviate concerns of residual temporal confounding. Results from our negative exposure control

analysis support this, with nonsignificant effect estimates. Lastly, we applied a validated phenotyping algorithm to distinguish heart failure subjects with reduced and preserved ejection fraction, a limitation of previous EHR-based epidemiologic studies of heart failure and environmental epidemiology studies of heart failure outcomes. These findings are relevant to persons in the study region yet would need to be evaluated separately in other geographies and populations with UNGD activity.

The study also had some limitations. First, we acknowledge that the use of ICD-9 codes to identify heart failure cases, while a sensitive measure, is less specific than other methods of case ascertainment, however we believe that our phenotype and severity analyses alleviate concerns about lack of specificity. Second, we did not have information on dietary intake and physical activity, and information on alcohol use was too often incomplete to use, however we did not believe that these factors would confound the association between UNGD activity and heart failure hospitalization. We did not have information on occupation, yet there is no evidence to suggest that current or past occupation would be highly correlated with UNGD activity metrics. Given the average age and general health status of these heart failure patients, it is unlikely that past occupational status accounts for observed associations or that these patients were employed in high exposure positions within the natural gas industry, if at all.

Conclusions and future directions

We observed significantly increased odds of hospitalization among heart failure subjects in relation to increasing UNGD activity for several phases, including pad preparation, stimulation, and production, with stronger associations among persons with more severe heart failure. These associations are plausible given environmental (e.g., air pollution (54), water contamination (55), noise (56), traffic (57)) and community (57,58) impacts of UNGD. Understanding how people living with heart failure are susceptible to environmental exposures, is especially important given the growing prevalence of heart failure and possibility that environmental factors play a role in clinical heart failure outcomes (10).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

The authors would like to thank Joseph DeWalle, Dione Mercer, and staff of the Geisinger Environmental Health Institute for their assistance in obtaining data essential for this study.

Financial support: This research was supported by funding from the National Institute of Environmental Health Sciences (NIEHS) training grant # 5T32ES007141-32 (T McAlexander) and NIEHS grant ES023675-01 (PI: B Schwartz). Additional support was provided by the Degenstein Foundation for the compiling of well data.

Abbreviations

BMI	body mass index
ED	emergency department

EHR	electronic health record
eMERGE	Electronic Medical Records and Genomics
HFpEF	heart failure with preserved ejection fraction
HF_rEF	heart failure with reduced ejection fraction
ICD	International Classification of Diseases
NDVI	normalized difference vegetation index
UNGD	unconventional natural gas development

References

1. Mozaffarian D, Benjamin EJ, Go AS et al. Heart Disease and Stroke Statistics—2016 Update: A Report From the American Heart Association. *Circulation* 2016;133:e38–e360.10.1161/cir.0000000000000350 [PubMed: 26673558]
2. Heart Failure Fact Sheet. Centers for Disease Control and Prevention. Accessed 10/1/15 http://www.cdc.gov/dhdsdp/data_statistics/fact_sheets/fs_heart_failure.htm
3. Savarese G, Lund LH. Global Public Health Burden of Heart Failure. *Card Fail Rev.* 2017;3:7–11.10.15420/cfr.2016:25:2 [PubMed: 28785469]
4. Heidenreich PA, Albert NM, Allen LA et al. Forecasting the Impact of Heart Failure in the United States: A Policy Statement From the American Heart Association. *Circ Heart Fail.* 2013;10:e001329.10.1161/HHF.0b013e318291329a
5. Heidenreich PA, Trogon JG, Khavjou OA et al. Forecasting the Future of Cardiovascular Disease in the United States. *Circulation* 2011;123:933. <http://circ.ahajournals.org/content/123/8/933.abstract> [PubMed: 21262990]
6. Jackson Sandra L, Tong X, King Raymond J, Loustalot F, Hong Y, Ritchey Matthew D. National Burden of Heart Failure Events in the United States, 2006 to 2014. *Circ Heart Fail.* 2018;11:e004873.10.1161/CIRCHEARTFAILURE.117.004873 [PubMed: 30562099]
7. von Lueder TG, Agewall S. The burden of heart failure in the general population: a clearer and more concerning picture. *J Thorac Dis.* 2018;10:S1934–S7.10.21037/jtd.2018.04.153 [PubMed: 30023084]
8. Zannad F Rising incidence of heart failure demands action. *Lancet.* 2018;391:518–9. 10.1016/S0140-6736(17)32873-8 [PubMed: 29174291]
9. Harjola V-P, Mullens W, Banaszewski M et al. Organ dysfunction, injury and failure in acute heart failure: from pathophysiology to diagnosis and management. A review on behalf of the Acute Heart Failure Committee of the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). *Eur J Heart Fail.* 2017;19:821–36.10.1002/ejhf.872 [PubMed: 28560717]
10. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. *Nat Rev Cardiol.* 2011;8:30–41.10.1038/nrcardio.2010.165 [PubMed: 21060326]
11. Bhatnagar A Environmental cardiology: studying mechanistic links between pollution and heart disease. *Circ Res.* 2006 9 29;99(7):692–705. [PubMed: 17008598]
12. Shah AS, Langrish JP, Nair H, McAllister DA, Hunter AL, Donaldson K, Newby DE, Mills NL. Global association of air pollution and heart failure: a systematic review and meta-analysis. *Lancet.* 2013 9 21;382(9897):1039–48. [PubMed: 23849322]
13. Sørensen M, Wendelboe Nielsen O, Sajadieh A, Kettel M, Tjønneland A, Overvad K, Raaschou-Nielsen O. Long-Term Exposure to Road Traffic Noise and Nitrogen Dioxide and Risk of Heart Failure: A Cohort Study. *Environ Health Perspect.* 2017 9 26;125(9):097021. [PubMed: 28953453]
14. Levin RK, Katz M, Saldiva PHN, et al. Increased hospitalizations for decompensated heart failure and acute myocardial infarction during mild winters: A seven-year experience in the public health

system of the largest city in Latin America. *PLoS One*. 2018 14;13(1):e0190733. [PubMed: 29300764]

15. Wellenius GA, Bateson TF, Mittleman MA, Schwartz J. Particulate Air Pollution and the Rate of Hospitalization for Congestive Heart Failure among Medicare Beneficiaries in Pittsburgh, Pennsylvania. *Am J Epidemiol*. 2005;161:1030–6.10.1093/aje/kwi135 [PubMed: 15901623]
16. Brook RD, Franklin B, Cascio W et al. Air Pollution and Cardiovascular Disease. *Circulation* 2004;109:2655, <http://circ.ahajournals.org/content/109/21/2655.abstract> [PubMed: 15173049]
17. PA DEP Oil & Gas Reporting Website Pennsylvania DEP. Accessed 7/2/16 <https://www.paoilandgasreporting.state.pa.us/publicreports/Modules/Welcome/Welcome.aspx>
18. Pennsylvania Shale Viewer.Fractracker Alliance. Accessed 12/3/19 <https://www.fractracker.org/map/us/pennsylvania/pa-shale-viewer/>
19. Boyle MD, Soneja S, Quirós-Alcalá L, et al. A pilot study to assess residential noise exposure near natural gas compressor stations. *PLoS One*. 2017 4 3;12(4):e0174310. [PubMed: 28369113]
20. McKenzie LM, Witter RZ, Newman LS, Adgate JL. Human health risk assessment of air emissions from development of unconventional natural gas resources. *Sci Total Environ*. 2012;424:79–87.10.1016/j.scitotenv.2012.02.018 [PubMed: 22444058]
21. Assessment of Risks from Unconventional Gas Well Development in the Marcellus Shale of Western MarylandMaryland DEP. Accessed 7/2/16 http://www.mde.state.md.us/programs/Land/mining/marcellus/Documents/A_Final_Draft_Cover_Ex_Sum_and_RA.pdf
22. Swarthout RF, Russo RS, Zhou Y et al. Impact of Marcellus Shale Natural Gas Development in Southwest Pennsylvania on Volatile Organic Compound Emissions and Regional Air Quality. *Environ Sci Technol*. 2015;49:3175–84.10.1021/es504315f [PubMed: 25594231]
23. Stallings-Smith S, Mease A, Johnson TM, Arikawa AY. Exploring the association between polycyclic aromatic hydrocarbons and diabetes among adults in the United States. *Environ Res*. 2018;166:588–94.10.1016/j.envres.2018.06.041 [PubMed: 29982146]
24. Fisher MP, Mayer A, Vollet K, Hill EL, Haynes EN. Psychosocial implications of unconventional natural gas development: Quality of life in Ohio’s Guernsey and Noble Counties. *Journal of Environmental Psychology* 2018;55:90–8.10.1016/j.jenvp.2017.12.008
25. Rabinowitz PM, Slizovskiy IB, Lamers V et al. Proximity to natural gas wells and reported health status: results of a household survey in Washington County, Pennsylvania. *Environ Health Perspect*. 2015;123:21–6.10.1289/ehp.1307732 [PubMed: 25204871]
26. Adgate JL, Goldstein BD, McKenzie LM. Potential Public Health Hazards, Exposures and Health Effects from Unconventional Natural Gas Development. *Environ Sci Technol*. 2014;48:8307–20.10.1021/es404621d [PubMed: 24564405]
27. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. *Annu Rev Public Health*. 2016;37:61–81.10.1146/annurev-publhealth-032315-021353 [PubMed: 26667605]
28. Wang MH, Shugart YY, Cole SR, Platz EA. A Simulation Study of Control Sampling Methods for Nested Case-Control Studies of Genetic and Molecular Biomarkers and Prostate Cancer Progression. *Cancer Epidemiol Biomarkers Prev*. 2009 3;18(3):706–11. [PubMed: 19258478]
29. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005 11;43(11):1130–9. [PubMed: 16224307]
30. Casey JA, Pollak J, Glymour MM, Mayeda ER, Hirsch AG, Schwartz BS. Measures of SES for Electronic Health Record-based Research. *Am J Prev Med*. 2018;54:430–9.10.1016/j.amepre.2017.10.004 [PubMed: 29241724]
31. Nau C, Schwartz BS, Bandeen-Roche K et al. Community socioeconomic deprivation and obesity trajectories in children using electronic health records. *Obesity (Silver Spring)* 2015;23:207–12.10.1002/oby.20903 [PubMed: 25324223]
32. Greene SJ, Butler J, Albert NM et al. Medical Therapy for Heart Failure With Reduced Ejection Fraction. *J Am Coll Cardiol*. 2018;72:35110.1016/j.jacc.2018.04.070 [PubMed: 30025570]
33. Hood SR, Giazzon AJ, Seamon G et al. Association Between Medication Adherence and the Outcomes of Heart Failure. *Pharmacotherapy: Pharmacotherapy*. 2018;38:539–45.10.1002/phar.2107 [PubMed: 29600819]

34. Tschöpe C, Birner C, Böhm M et al. Heart failure with preserved ejection fraction: current management and future strategies. *Clin Res Cardiol*. 2018;107:1–19.10.1007/s00392-017-1170-6
35. van Ginkel JR, Linting M, Rippe RCA, van der Voort A. Rebutting Existing Misconceptions About Multiple Imputation as a Method for Handling Missing Data. *J Pers Assess*. 2020 May-Jun;102(3):297–308. [PubMed: 30657714]
36. Bielinski SJ, Pathak J, Carrell DS et al. A Robust e-Epidemiology Tool in Phenotyping Heart Failure with Differentiation for Preserved and Reduced Ejection Fraction: the Electronic Medical Records and Genomics (eMERGE) Network. 2015,
37. Bielinski SJ. Heart Failure (HF) with Differentiation between Preserved and Reduced Ejection Fraction In: *Clinic. M*, editor Phe KB, 2013, <https://phekb.org/phenotype/147>
38. Casey JA, Savitz DA, Rasmussen SG, Ogburn EL, Pollak J, Mercer DG, Schwartz BS. Unconventional Natural Gas Development and Birth Outcomes in Pennsylvania, USA. *Epidemiology*. 2016 3;27(2):163–72. [PubMed: 26426945]
39. Koehler K, Ellis JH, Casey JA et al. Exposure Assessment Using Secondary Data Sources in Unconventional Natural Gas Development and Health Studies. *Environ Sci Technol*. 2018;52:6061–9.10.1021/acs.est.8b00507 [PubMed: 29697245]
40. Liu AY, Curriero FC, Glass TA, Stewart WF, Schwartz BS. The contextual influence of coal abandoned mine lands in communities and type 2 diabetes in Pennsylvania. *Health Place*. 2013 7;22:115–22. [PubMed: 23689181]
41. Schwartz BS, Stewart WF, Godby S et al. Body Mass Index and the Built and Social Environments in Children and Adolescents Using Electronic Health Records. *Am J Prev Med*. 2011;41:e17–e28.10.1016/j.amepre.2011.06.038 [PubMed: 21961475]
42. American Community Survey.United States Census Bureau Accessed 10/1/15 <https://www.census.gov/programs-surveys/acs/>
43. Nau C, Schwartz BS, Bandeen-Roche K et al. Community socioeconomic deprivation and obesity trajectories in children using electronic health records. *Obesity (Silver Spring)* 2015;23:207–12.10.1002/oby.20903 [PubMed: 25324223]
44. Tustin AW, Hirsch AG, Rasmussen SG, Casey JA, Bandeen-Roche K, Schwartz BS. Associations between Unconventional Natural Gas Development and Nasal and Sinus, Migraine Headache, and Fatigue Symptoms in Pennsylvania. *Environ Health Perspect*. 201610.1289/ehp281
45. Casey J,A, Ogburn E,L, Rasmussen S,G et al. Predictors of Indoor Radon Concentrations in Pennsylvania, 1989–2013. *Environ Health Perspect*. 2015;123:1130–7.10.1289/ehp.1409014 [PubMed: 25856050]
46. Rasmussen SG, Ogburn EL, McCormack M et al. Association Between Unconventional Natural Gas Development in the Marcellus Shale and Asthma Exacerbations.. *JAMA Intern Med*. 2016LID - 10.1001/jamainternmed.2016.2436 [doi]
47. Ward-Caviness Cavin K, Weaver Anne M, Buranosky M et al. Associations Between Long-Term Fine Particulate Matter Exposure and Mortality in Heart Failure Patients. *J Am Heart Assoc*. 2020 3 17;9(6):e012517. [PubMed: 32172639]
48. Lo AX, Donnelly JP, Durant RW et al. Emergency Departments: Racial Disparities in Hospitalizations for Heart Failure. *Am J Prev Med*. 2018 11;55(5 Suppl 1):S31–S39. [PubMed: 30670199]
49. Weiskopf MG, Tchetgen Tchetgen EJ, Raz R. Commentary: On the Use of Imperfect Negative Control Exposures in Epidemiologic Studies. *Epidemiology*. 2016 5;27(3):365–7. [PubMed: 27035687]
50. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010 5;21(3):383–8. doi: 10.1097/EDE.0b013e3181d61eeb Erratum in: *Epidemiology*. 2010 Jul;21(4):589. [PubMed: 20335814]
51. Union of Concerned Scientists.Toward an Evidence-Based Fracking Debate: Science, Democracy, and Community Right to Know in Unconventional Oil and Gas Development. Accessed 10/1/15 <http://www.ucsusa.org/center-for-science-and-democracy/toward-an-evidence-based-fracking-debate.html#.VxAzoxMrKuU>

52. Vengosh A, Jackson RB, Warner N, Darrah TH, Kondash A. A Critical Review of the Risks to Water Resources from Unconventional Shale Gas Development and Hydraulic Fracturing in the United States. *Environ Sci Technol*. 2014;48:8334–48.10.1021/es405118y [PubMed: 24606408]
53. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of heart failure diagnoses in administrative databases: a systematic review and meta-analysis. *PloS one* 2014;9:e104519–e.10.1371/journal.pone.0104519 [PubMed: 25126761]
54. Srebotnjak T, Rotkin-Ellman M. Fracking Fumes: Air pollution from hydraulic fracturing threatens public health and communities. Natural Resources Defense Council. Accessed 9/25/20 <https://www.nrdc.org/sites/default/files/fracking-air-pollution-IB.pdf>
55. Osborn SG, Vengosh A, Warner NR, Jackson RB. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proc Natl Acad Sci U S A*. 2011 5 17;108(20):8172–6. doi: 10.1073/pnas.1100682108. [PubMed: 21555547]
56. Hays J, Shonkoff SB. Toward an Understanding of the Environmental and Public Health Impacts of Unconventional Natural Gas Development: A Categorical Assessment of the Peer-Reviewed Scientific Literature, 2009–2015. *PLoS One*. 2016 4 20;11(4):e0154164. [PubMed: 27096432]
57. Increased traffic accident rates associated with shale gas drilling in Pennsylvania. *Accid Anal Prev*. 2015 1;74:203–9. [PubMed: 25463961]
58. Perry SL. Development, Land Use, and Collective Trauma: The Marcellus Shale Gas Boom in Rural Pennsylvania. *CAFÉ*. 2012;34: 81–92.

Perspectives

Competency in Systems-Based Practice:

Exposure to unconventional natural gas development (UNGD) activities, including hydraulic fracturing (“fracking”) increases the risk of hospitalization among patients with heart failure.

Translational Outlook:

Additional research is needed to understand the pathophysiological mechanisms through which UNGD can exacerbate heart failure.

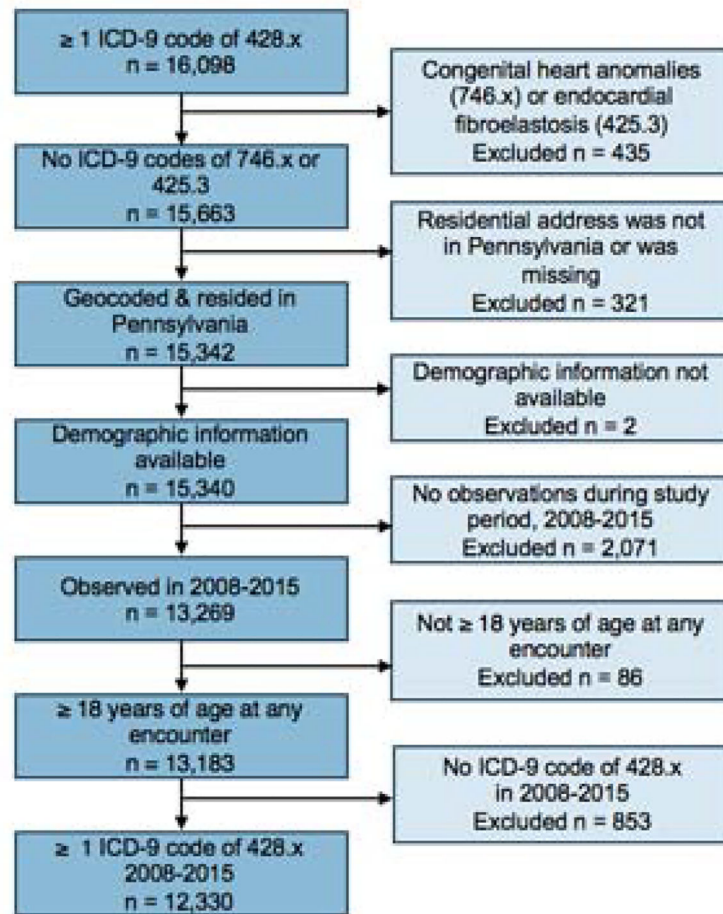


Figure 1. Schematic of inclusion and exclusion of study subjects.

This flowchart details the inclusion and exclusion criteria used for identifying this study population from the electronic health record.

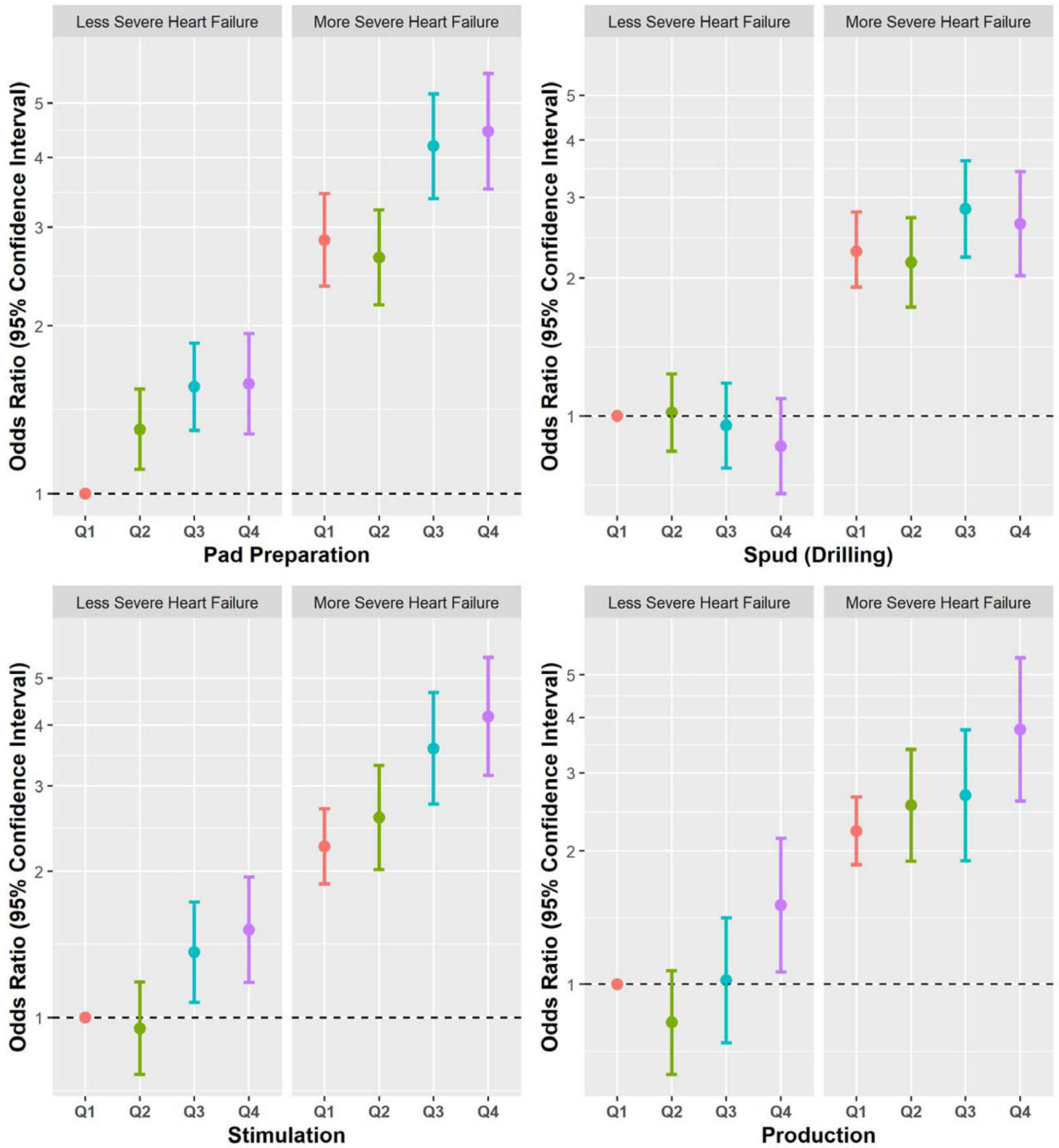


Figure 2. Associations (OR and 95% CI) of quartiles of UNGD activity metrics with hospitalization by heart failure severity.

The “More Severe Heart Failure” group included HFpEF and HFrEF; the “Less Severe Heart Failure” group included subjects with the electronic medical records and genomics (eMERGE) algorithm not applied and those with no discernable phenotype by eMERGE. Reference groups included less severe patients in the first quartile of each unconventional natural gas development (UNGD) metric, and colors denote quartiles (Q1= red, Q2 = green, Q3 = blue, Q4 = purple) of respective UNGD activity metrics (pad preparation, spud

[drilling], stimulation [i.e., “fracking”], and production). Effect modification was present for three of the four phases

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

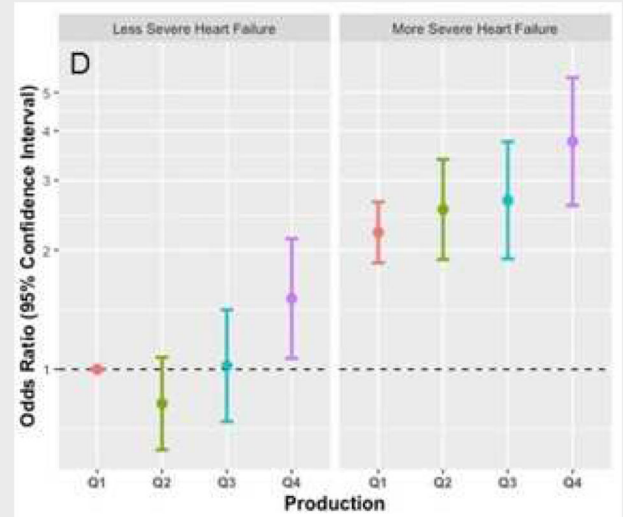
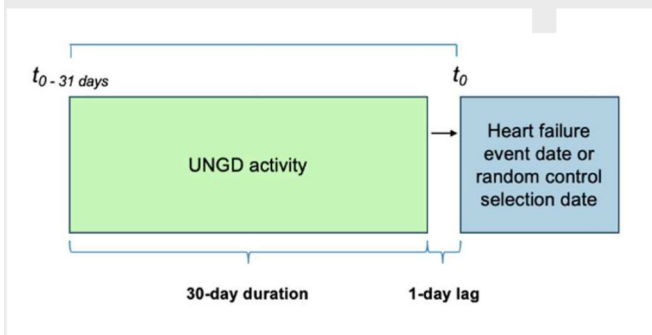
**Primary Exposure of Interest:
Unconventional Natural Gas
Development (UNGD) Activity**

- By phase of development: pad development, drilling, stimulation (“fracking”), and production
- Included 9669 wells statewide

Primary Findings

- Three of four phases of UNGD activity were associated with increased odds of heart failure hospitalization
- Associations of UNGD activity with heart failure hospitalization were stronger in persons with more severe heart failure

UNGD Activity Assignment



Central Illustration.

Associations between unconventional natural gas development activity and hospitalization among heart failure patients in Pennsylvania.

Table 1.

Selected patient characteristics by case and control status, at time of case event or control selection date

	Never a case	Ever a case	
	n = 3215	n = 5839	p-value*
Total patients n = 9054			
Sex, n (%)			
Male	1659 (51.6)	3074 (52.7)	
Female	1556 (48.4)	2765 (47.4)	0.34
Age at hospitalization or control selection date, years, mean (SD)	71.0 (12.6)	71.1 (12.7)	0.75
Race/ethnicity, n (%)			
White	3127 (97.3)	5688 (97.4%)	
Black	46 (1.4)	74 (1.3%)	
Hispanic	28 (0.9)	60 (1.0%)	
Other	10 (0.3)	17 (0.3%)	
Missing	4 (0.1)	0 (0%)	0.08
Smoking status at event, n (%)			
Current	358 (11.1)	760 (13.0)	
Former	1539 (47.9)	2690 (46.1)	
Never	1318 (41.0)	2389 (40.9)	0.03
Community type, n (%)			
Borough	991 (30.8%)	1834 (31.4%)	
Township	1926 (59.9%)	3248 (55.6%)	
Census tract (city)	298 (9.3%)	757 (12.9%)	< 0.01
CSD[†], SD units, quartiles			
1 (-7.5, < -2.6)	615 (19.1%)	1081 (18.5%)	
2 (-2.6, < -0.5)	823 (25.6%)	1419 (24.3%)	
3 (-0.5, < 2.3)	904 (28.1%)	1775 (30.4%)	
4 (2.3, < 22.6)	873 (27.2%)	1564 (26.8%)	0.13

	Never a case	Ever a case	
	n = 3215	n = 5839	p-value*
Total patients n = 9054			
Distance to major road [‡] (meters), mean (SD)	2908 (4,160)	2703 (4,282)	0.03
Distance to minor road [‡] (meters), mean (SD)	1784 (2599)	1431 (2136)	<0.01
Receipt of Medical Assistance [§] , n (%)	368 (11.5)	681 (11.7)	0.76
Body mass index (BMI) at event, kg/m ² , mean (SD)	31.9 (7.7)	31.5 (8.8)	0.51
Duration of contact with health system ² , days, mean (SD)	3995 (1395)	3683 (1530)	<0.01
Medication use [#] , by class, n (%)			
Antihypertensive	1418 (44.1)	2324 (39.8)	<0.01
Antihyperlipidemic	1503 (46.8)	2615 (44.8)	0.07
Anticoagulant	625 (19.4)	1049 (18.0)	0.08
Chronic obstructive pulmonary disease (COPD), n (%)	421 (13.1)	1106 (18.9)	<0.01
Coronary artery disease, n (%)	492 (15.3)	1220 (20.9)	<0.01
Hypertension, n (%)	1801 (56.0)	3962 (67.9)	<0.01
Myocardial infarction, n (%)	212 (6.59)	433 (7.42)	0.15
Valve disorder, n (%)	496 (15.4)	1127 (19.3)	<0.01
Diabetes, n (%)	989 (30.8)	2356 (40.3)	<0.01
Chronic kidney disease, n (%)	610 (26.4)	1519 (29.6)	<0.01
Charlson Index of morbidity ^{***} , mean (SD)	8.26 (3.18)	8.76 (3.36)	<0.01
Patient status at end of study, n (%)			
Alive	2514 (78.2%)	3487 (59.7%)	
Deceased	701 (21.8%)	2352 (40.3%)	<0.01

* p-value obtained from chi² tests comparing selected variable in cases and controls for categorical or binary variables; ANOVA F-test for continuous variables

⁷Community socioeconomic deprivation (CSD) was calculated based on US Census indicators; further information is detailed in the text

⁸Major and minor roads were identified from the Pennsylvania Department of Transportation (DOT) databases; distance from subject's residential address to these roads was calculated in meters using the Generate Near Table tool function in ArcGIS 10.4.

[§]Medical Assistance, a surrogate for family socioeconomic status, was calculated based on health insurance status at the time of encounters as previously reported

[?]Days from first to most recent (i.e., case event or control selection date) time a subject was observed in the HER

[#]Relevant medication classes were identified based on the dates of physician orders

^{**}A composite measure of overall morbidity; definition described in text

Table 2.

Selected patient characteristics by phenotype status, at time of randomly selected case event or control selection date

	Total patients n = 9054	Not phenotyped n = 5702	HFpEF n = 1739	HFpEF n = 1613	p-value*
Sex, n (%)					
Male	2978 (52.2)	1107 (63.7)	648 (40.2)		< 0.01
Female	2724 (47.8)	632 (36.3)	965 (59.8)		
Race/ethnicity, n (%)					
White	5527 (96.9)	1705 (98.0)	1583 (98.1)		
Black	95 (1.7)	12 (0.7)	13 (0.8)		
Hispanic	59 (1.0)	16 (0.9)	13 (0.8)		
Other	18 (0.3)	5 (0.3)	4 (0.3)		
Missing	3 (0.05)	1 (0.06)	0 (0.0)		0.05
Age at hospitalization or control selection date, years, mean (SD)	70.5 (13.0)	70.3 (12.5)	74.0 (11.2)		< 0.01
Age category at first event, n (%) years					
> 18–30	49 (0.9)	11 (0.6)	1 (0.06)		
> 30–40	73 (1.3)	26 (1.5)	10 (0.6)		
> 40–50	281 (4.9)	80 (4.6)	36 (2.2)		
> 50–60	768 (13.5)	234 (13.6)	152 (9.4)		
> 60–70	1337 (23.5)	402 (23.1)	319 (19.8)		
> 70–80	1596 (28.0)	524 (30.1)	530 (32.9)		
> 80–90	1526 (26.8)	449 (25.8)	536 (33.2)		
> 90–100	72 (1.3)	10 (0.6)	29 (1.8)		< 0.01
Community type, n (%)					
Borough	1735 (30.4)	578 (33.2)	512 (31.7)		
Township	3245 (56.9)	998 (57.4)	931 (57.7)		
Census tract (city)	722 (12.7)	163 (9.4)	170 (10.5)		< 0.01
Community socioeconomic deprivation (CSD) [†], SD units, quartiles					
I (-7.5, < -2.6)	1046 (18.3)	338 (19.4)	312 (19.3)		

	Total patients n = 9054	Not phenotyped n = 5702	HFpEF n = 1739	HFpEF n = 1613	p-value*
	2 (-2.6, < -0.5)	1415 (24.8)	436 (25.1)	391 (24.2)	
	3 (-0.5, < 2.3)	1672 (29.3)	513 (29.5)	494 (30.6)	
	4 (2.3, < 22.6)	1569 (27.5)	452 (26.0)	416 (25.8)	0.64
Patient status at end of study, n (%)					
Alive		3969 (69.6)	1058 (60.8)	974 (60.4)	
Deceased		1733 (30.4)	681 (39.2)	639 (39.6)	< 0.01
Distance to major road[‡], meters, mean (SD)					
		2794 (4228)	2877 (4295)	2603 (4219)	0.15
Distance to minor road[‡], meters, mean (SD)					
		1600 (2400)	1536 (2271)	1423 (2049)	0.02
Distance to hospital/clinic, meters, mean (SD)					
		6914 (8484)	6531 (8112)	6111 (7778)	< 0.01
Smoking status at event, n (%)					
Current		721 (12.6)	247 (14.2)	150 (9.3)	
Former		2624 (46.0)	856 (49.2)	749 (46.4)	
Never		2357 (41.3)	636 (36.6)	714 (44.3)	< 0.01
Receipt of Medical Assistance[§], n (%)					
		679 (11.9)	198 (11.4)	172 (10.7)	0.37
Body mass index (BMI) at event, kg/m², mean (SD)					
		31.6 (8.4)	30.4 (7.5)	32.9 (9.0)	< 0.01
Time since first heart failure diagnosis[‡], days, mean (SD)					
		855 (1148)	754 (1099)	670 (1020)	< 0.01
Medication use[#], by class, n (%)					
Antihypertensive		2227 (39.1)	803 (46.2)	712 (44.1)	< 0.01
Antihyperlipidemic		2467 (43.3)	831 (47.8)	820 (50.8)	< 0.01
Anticoagulant		967 (17.0)	340 (19.6)	367 (22.8)	< 0.01
Chronic obstructive pulmonary disease, n (%)					
		1268 (22.2)	354 (20.4)	430 (26.7)	< 0.01
Coronary artery disease, n (%)					
		1182 (20.7)	520 (29.9)	421 (26.1)	< 0.01
Hypertension, n (%)					
		4284 (75.1)	1331 (76.5)	1372 (85.1)	< 0.01

Total patients n = 9054	Not phenotyped n = 5702	HFpEF n = 1739	HFpEF n = 1613	p-value*
Myocardial infarction, n (%)	688 (12.1)	340 (19.6)	143 (8.9)	< 0.01
Valve disorder, n (%)	1081 (19.0)	425 (24.4)	481 (29.8)	< 0.01
Diabetes, n (%)	2365 (41.5)	797 (45.8)	783 (48.5)	< 0.01
Chronic kidney disease, n (%)	1479 (25.9)	528 (30.4)	571 (35.4)	< 0.01
Charlson index of morbidity ^{***} , mean (SD)	8.3 (3.3)	8.7 (3.4)	9.4 (3.2)	< 0.01

* p-value obtained from either χ^2 tests (for categorical variables) or analysis of variance (ANOVA) F-test (for continuous variables), comparing events for subjects who had an assigned phenotype (heart failure with reduced ejection fraction [HFpEF] or preserved ejection fraction [HFpEF]) vs. those who did not have phenotype information

[†] Community socioeconomic deprivation (CSD) was calculated based on US Census indicators; further information is detailed in the text

[‡] Major & minor roads were identified from the Federal Highway Administration databases; distance from subject's residential address to these roads was calculated in meters

[§] Medical Assistance, a surrogate for family socioeconomic status, was calculated based on health insurance status at the time of encounters

[?] Days from first heart failure diagnosis to the date of case or control event.

[#] Relevant medication classes were identified based on the dates of physician orders

^{***} A composite measure of overall morbidity; definition described in text

Table 3.

Associations of UNGD activity metrics, by phase and quartile (Q), with hospitalization for heart failure, from models with increasing covariate adjustment

UNGD Metric	Model 1* Odds ratio (95% CI)	Model 2 [†] Odds ratio (95% CI)	Model 3 [‡] Odds ratio (95% CI)	Model 4 [§] Odds ratio (95% CI)	Model 5 [¶] Odds ratio (95% CI)
Pad preparation, 1/m²					
Q1 (4.0×10^{-9} , < 3.3×10^{-8})	(ref)	(ref)	(ref)	(ref)	(ref)
Q2 (3.3×10^{-8} , < 6.2×10^{-8})	1.07 (0.92, 1.25)	1.23 (1.04, 1.46)	1.03 (0.88, 1.19)	1.16 (0.98, 1.37)	1.19 (1.01, 1.40)
Q3 (6.2×10^{-8} , < 1.2×10^{-7})	1.39 (1.19, 1.62)	1.75 (1.44, 2.13)	1.34 (1.15, 1.57)	1.65 (1.36, 2.00)	1.63 (1.35, 1.97)
Q4 (1.2×10^{-7} , < 2.5×10^{-5})	1.29 (1.10, 1.52)	1.84 (1.45, 2.33)	1.24 (1.06, 1.46)	1.70 (1.34, 2.15)	1.70 (1.35, 2.13)
Spud (drilling), 1/m²					
Q1 (1.9×10^{-10} , < 2.6×10^{-9})	(ref)	(ref)	(ref)	(ref)	(ref)
Q2 (2.6×10^{-9} , < 3.8×10^{-9})	1.26 (1.08, 1.47)	0.94 (0.76, 1.16)	1.23 (1.06, 1.43)	1.01 (0.82, 1.25)	1.01 (0.82, 1.25)
Q3 (3.8×10^{-9} , < 4.7×10^{-8})	1.26 (1.09, 1.47)	0.90 (0.71, 1.13)	1.27 (1.09, 1.48)	1.04 (0.82, 1.31)	1.07 (0.85, 1.35)
Q4 (4.7×10^{-9} , < 1.7×10^{-8})	0.96 (0.82, 1.13)	0.64 (0.49, 0.83)	1.16 (0.99, 1.37)	0.93 (0.71, 1.21)	0.97 (0.75, 1.27)
Stimulation, m/m²					
Q1 (2.5×10^{-6} , < 2.7×10^{-4})	(ref)	(ref)	(ref)	(ref)	(ref)
Q2 (2.7×10^{-4} , < 5.6×10^{-4})	0.96 (0.82, 1.12)	1.03 (0.80, 1.31)	0.99 (0.85, 1.16)	1.00 (0.78, 1.28)	1.03 (0.81, 1.31)
Q3 (5.6×10^{-4} , < 9.7×10^{-4})	1.45 (1.24, 1.70)	1.67 (1.27, 2.20)	1.41 (1.20, 1.65)	1.50 (1.14, 1.98)	1.56 (1.19, 2.04)
Q4 (9.7×10^{-4} , < 0.2)	1.65 (1.39, 1.96)	1.98 (1.47, 2.67)	1.62 (1.37, 1.92)	1.78 (1.32, 2.40)	1.80 (1.35, 2.40)
Production, m³/m²					
Q1 (2.2×10^{-6} , < 0.002)	(ref)	(ref)	(ref)	(ref)	(ref)
Q2 (0.002, < 0.02)	1.01 (0.86, 1.18)	0.89 (0.65, 1.21)	1.07 (0.92, 1.26)	0.83 (0.61, 1.14)	0.87 (0.64, 1.19)
Q3 (0.02, < 0.03)	1.19 (1.01, 1.40)	1.34 (0.92, 1.96)	1.15 (0.98, 1.34)	1.00 (0.69, 1.47)	1.10 (0.75, 1.60)
Q4 (0.03, < 16.5)	1.72 (1.44, 2.06)	2.17 (1.42, 3.30)	1.66 (1.40, 1.98)	1.54 (1.01, 2.34)	1.62 (1.07, 2.45)

* Model 1: [binary indicators] sex, race/ethnicity (white vs. non-white), receipt of Medical Assistance (ever/never), smoking status (ever/never), body mass index (BMI, kg/m²) (centered and centered-squared term), age category

[†] Model 2: Model 1 + year

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

[‡]Model 3: Model 1+ region. Regions were defined by county of residence as northeast (reference), southeast, central, southwest, and northwest.

[§]Model 4: Model 1+ year and region

[?]Model 5: Model 1+ year, region, season (winter [reference], spring, summer, fall), distance to hospital/clinic, and contact time (date of case or control encounter minus the date of the first encounter in medical record, in days)

Adjusted associations (odds ratio [OR], 95% confidence interval [CI]) of UNGD activity metrics with hospitalization by heart failure phenotype strata

Table 4.

UNGD Metric	Quartile (Range)	Phenotype Odds ratio (95% CI)*		p-value for interaction [†]
		HFpEF	HFrEF	
Pad preparation, 1/m ²	1 (4.0 × 10 ⁻⁹ , < 3.3 × 10 ⁻⁸)	Reference	0.95 (0.70, 1.29)	
	2 (3.3 × 10 ⁻⁸ , < 6.2 × 10 ⁻⁸)	0.70 (0.52, 0.95)	1.17 (0.95, 1.61)	
	3 (6.2 × 10 ⁻⁸ , < 1.2 × 10 ⁻⁷)	1.36 (0.98, 1.88)	1.52 (1.10, 2.10)	
	4 (1.2 × 10 ⁻⁷ , < 2.5 × 10 ⁻⁵)	1.56 (1.11, 2.19)	1.52 (1.09, 2.13)	0.05
Spud (drilling), 1/m ²	1 (1.9 × 10 ⁻¹⁰ , < 2.6 × 10 ⁻⁹)	Reference	1.28 (0.96, 1.71)	
	2 (2.6 × 10 ⁻⁹ , < 3.8 × 10 ⁻⁹)	0.99 (0.72, 1.37)	1.13 (0.81, 1.57)	
	3 (3.8 × 10 ⁻⁹ , < 4.7 × 10 ⁻⁸)	1.47 (1.05, 2.05)	1.31 (0.94, 1.84)	
	4 (4.7 × 10 ⁻⁹ , < 1.7 × 10 ⁻⁸)	1.09 (0.77, 1.55)	1.53 (1.08, 2.17)	0.17
Stimulation, m/m ²	1 (2.5 × 10 ⁻⁶ , < 2.7 × 10 ⁻⁴)	Reference	1.40 (1.05, 1.86)	
	2 (2.7 × 10 ⁻⁴ , < 5.6 × 10 ⁻⁴)	1.21 (0.86, 1.71)	1.47 (1.04, 2.06)	
	3 (5.6 × 10 ⁻⁴ , < 9.7 × 10 ⁻⁴)	1.72 (1.21, 2.44)	2.05 (1.43, 2.92)	
	4 (9.7 × 10 ⁻⁴ , < 0.2)	2.25 (1.56, 3.25)	2.09 (1.44, 3.03)	0.27
Production, m ³ /m ²	1 (2.2 × 10 ⁻⁶ , < 0.002)	Reference	1.38 (1.04, 1.83)	
	2 (0.002, < 0.02)	1.32 (0.91, 1.92)	1.35 (0.93, 1.96)	
	3 (0.02, < 0.03)	1.20 (0.79, 1.81)	1.50 (1.03, 2.34)	
	4 (0.03, < 16.5)	1.96 (1.26, 3.04)	1.94 (1.24, 3.04)	0.33

* Multilevel logistic regression models adjusted for: sex (male vs. female), race/ethnicity (white vs. nonwhite), receipt of Medical Assistance (ever vs. never received prior to case event or control encounter), smoking status (ever vs. never), body mass index (BMI, as a centered and centered squared term), age category at case event or control encounter, days since first heart failure diagnosis code (centered), distance to nearest hospital or clinic (centered), season, region, year, and three phenotype indicators: heart failure with reduced ejection fraction (yes vs. no), eMERGE algorithm applied, but unable to be phenotyped (yes vs. no), and eMERGE algorithm not applied (yes vs. no), and cross-products between each phenotype indicator and UNGD metric.

[†] p-value obtained from chi² tests for the cross-products between quartiles of UNGD activity and HFrEF phenotype indicator variable