



HEI Research Report 233

Hydrocarbon Extraction and Risk to Groundwater in Pennsylvania: Part 2. An Analysis of Water Quality in the Oldest Hydrocarbon Basin in the United States

Jennifer Baka et al.

Additional Materials 5a. Survey Statistics Code

Correspondence may be addressed to Dr. Jennifer Baka, Department of Geography, 302 Walker Building, University Park, PA 16802; email: jeb525@psu.edu.

Although this report was produced with partial funding by the United States Environmental Protection Agency under Contract No. 68HERC19D0010 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and may not reflect the views of the Agency; thus, no official endorsement by the Agency should be inferred. This report also has not been reviewed by private party institutions, including those that support HEI Energy, and may not reflect the views or policies of these parties; thus, no endorsement by them should be inferred.

Additional Materials 5a. Survey Statistics Code

```
libname BitzerD 'C:\Users\tmurphy11\OneDrive - Penn State  
Health\Migration\BiostatConsults\2025\Bitzer\data';
```

```
options ls=75 nofmterr;
```

```
*****;  
*****;  
*****;  
*****;  
*****;
```

```
*** import and examine BL1 data ;
```

```
proc import datafile = 'C:\Users\tmurphy11\OneDrive - Penn State  
Health\Migration\BiostatConsults\2025\Bitzer\data\prunedRedCAPdataDec1124temFeb2825.  
csv'
```

```
    out=Bitzer  
    dbms=csv  
    replace;  
        *delimiter=',';  
        getnames=yes;  
run;    *** 394 obs and 38 vars ;
```

```
proc print data=Bitzer (obs = 10) ;  
run;    *** some lines of data with 'do not use' ;
```

```
proc contents data=Bitzer;  
title 'Contents of Bitzer Data';  
run;    *** rename titles in CSV as shorter ;
```

```
*****;  
*****;  
*****;  
*****;  
*****;
```

```
proc freq data=Bitzer;  
tables  
site;  
run;    *** 174 rows with non-missing site, 220 with missing vaues;
```

```

*** throw out rows where site missing;

data    BitzerSites;
set      Bitzer;
where site ne . ;
run;    *** 174 obs and 38 vars ;

proc sort data=BitzerSites out=sortsites;
by site;
run;

data    uniqueSites;
set      sortSites;
by site;
if last.site;
run;    *** 90 obs and 38 vars ;

proc print data=uniqueSites (obs = 20);
run;

ods rtf file = 'C:\Users\tmurphy11\OneDrive - Penn State
Health\Migration\BiostatConsults\2025\Bitzer\data\ContentsUniqueSitesMar0625.rtf'
style=journal;
ods graphics on;
proc contents data=uniqueSites;
title 'Contents of UniqueSites Data (n=90)';
run;
ods graphics off;
ods rtf close;

data    BitzerD.SitesMar0625;
set      uniqueSites;
run;    *** 90 obs and 38 sites;

*** export as CSV and send to Zach for him to suggest some analytical goals;

proc export data=BitzerD.SitesMar0625
    outfile="C:\Users\tmurphy11\OneDrive - Penn State
Health\Migration\BiostatConsults\2025\Bitzer\data\SitesMar0625.csv"
    dbms=csv
    replace;
run;

```

```
*****;  
*****;  
*****;  
*****;  
*****;  
*****;
```

```
*** create desc stats of vars requested by Zach Bitzer;
```

```
ods rtf file = 'C:\Users\tmurphy11\OneDrive - Penn State  
Health\Migration\BiostatConsults\2025\Bitzer\data\DescStatsUniqueSitesMar0625.rtf'  
style=journal;  
ods graphics on;
```

```
proc means data=BitzerD.SitesMar0625 n nmiss min max mean median ;  
title 'Descr Stats of Continuous Variables in Bitzer Data';  
var  
yrH0meBuilt  
wellDepthFeet  
distNearOilGasWellPad  
distNearImpPond  
distNearGasStat  
;  
run;
```

```
proc freq data=BitzerD.SitesMar0625 ;  
title 'Descr Stats of Categorical Variables in Bitzer Data';  
tables  
sourceDkWater  
whyNotWellDrink  
BWforDrkWater  
growFrVeg  
waterFrVeg  
raiseFoodAnim  
animalsWellWater  
animalsCropsFromProp  
;  
run;
```

```
ods graphics off;  
ods rtf close;
```

*****;
*****;
*****;
*****;
*****;

*** here Thur Mar 6, 2025;

*****;
*****;
*****;
*****;
*****;

***** das Ende
*****;