Supporting Information Appendix for

Children Drinking Private Well Water Have Higher Blood Lead Than Those with City Water

Jacqueline MacDonald Gibson, Michael Fisher, Allison Clonch, John M. MacDonald, Philip J. Cook
Correspondence to: jacmgibs@iu.edu

This PDF file includes:

- Figs. S1 to S2
- Tables S1 to S5
- Stata code for building and analyzing curated data set

www.pnas.org/cgi/doi/10.1073/pnas.2002729117

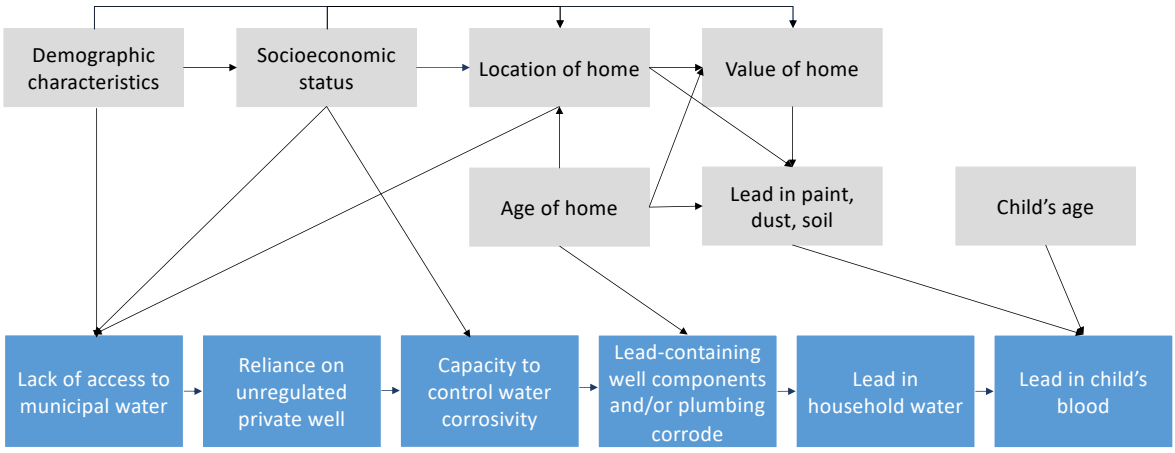

Fig. S1.

Conceptual model of children's exposure to Pb in drinking water in houses relying on unregulated private wells.

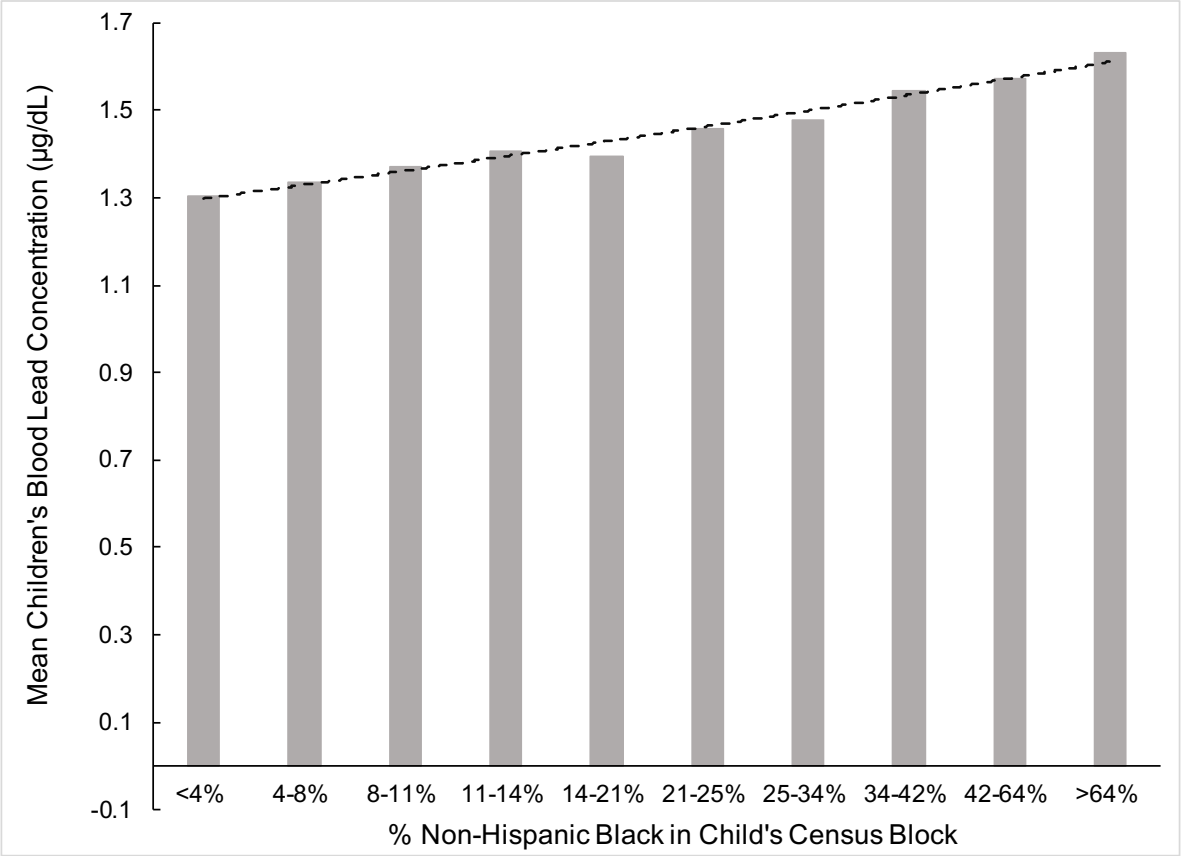

Fig. S2.
Blood lead concentration by % non-Hispanic black in the child's census block.

Independent Variable	Coefficient	Std. Err.	*t*	*p>t*	Lower 95% CI	Upper 95% CI
Water source (0=community water, 1=private well) (*n*=51,691)	0.0797	0.00619	12.9	<0.001	0.0675	0.0918
Male biological sex (*n*=58,330)	0.00656	0.00445	1.48	0.14	-0.00215	0.0153
Home value (natural log of total $ value) (*n*=55,408)	-0.141	0.00368	-38.5	<0.001	-0.149	-0.134
Born in extraterritorial jurisdiction (0=no, 1=yes) (*n*=54,678)	0.0977	0.00924	10.6	<0.001	0.0800	0.116
Blood draw year (natural log of years since 1990) (*n*=59,483)	-1.058	0.00825	-128	<0.001	-1.074	-1.042
Home area (sq. ft.) (*n*=59,483)	-1.32E-07	8.67E-08	-1.52	0.129	-3.02E-07	3.85E-08
Blood draw type (venous vs. capillary) (*n*=56,322)	-0.012	0.00750	-1.55	0.121	-0.0263	0.00307
African American population proportion (*n*=59,466)	0.316	0.0101	31.5	<0.001	0.300	0.336
Hispanic population proportion (*n*=59,466)	0.500	0.0252	19.9	<0.001	0.451	0.549
Median income (natural log) (*n*=59,461)	-0.190	0.00508	-37.3	<0.001	-0.200	-0.180
Age (months; reference category 0-8 months) (*n*=59,483)						
9-14	-0.011	0.0245	-0.45	0.652	-0.0591	0.0369
15-19	0.141	0.0258	5.46	<0.001	0.0904	0.192
20-29	0.061	0.0250	2.43	0.015	0.0117	0.110
30+	0.180	0.0252	7.14	<0.001	0.130	0.229
Year house was built (ref. period: before 1950) (*n*=58,091)						
1950-1977	-0.078	0.0102	-7.68	<0.001	-0.0982	-0.0583
1978-1987	-0.129	0.0108	-12.0	<0.001	-0.150	-0.108
1988-1997	-0.153	0.0104	-14.7	<0.001	-0.174	-0.133
1998-2002	-0.161	0.0104	-15.5	<0.001	-0.182	-0.141
2003-2007	-0.268	0.0104	-25.9	<0.001	-0.289	-0.248
2008-2017	-0.391	0.0119	-32.7	<0.001	-0.414	-0.367
Season (reference: winter)						
Spring	0.00247	0.00447	0.393	0.694	-0.00982	0.0148
Summer	0.0160	0.00616	2.59	0.010	0.00389	0.0280
Fall	-0.00402	0.00635	-0.634	0.526	-0.0165	0.00842

Table S1.

Univariable ordinary least squares regressions of log blood lead versus selected independent variables. A test of the interactions between private well and year of construction was not statistically significant (Chi-square=3.46, df=6, *p*=0.748).

Independent Variable	Exponentiated Coefficient	Std Err.	*Z*	*p>Z*	Lower 95% CI	Upper 95% CI
Private well (reference category: community water supply)	1.20	0.0266	8.13	<0.001	1.15	1.25
Male biological sex	1.05	0.0119	3.83	<0.001	1.02	1.07
Home value (natural log of total $)	0.77	0.0150	-13.3	<0.001	0.745	0.804
Born in extraterritorial jurisdiction (ETJ) (reference category: not born in ETJ)	1.09	0.0359	2.53	0.012	1.02	1.16
Blood draw year-2002 (natural log)	0.0671	0.00236	-76.8	<0.001	0.0626	0.0719
Home area (sq. ft.)	1.00	1.16E-06	8.89	<0.001	1.00	1.00
Blood draw type (venous=0; capillary=1)	1.10	0.0282	3.51	<0.001	1.04	1.15
African American population proportion	1.29	0.0664	4.99	<0.001	1.17	1.43
Hispanic population proportion	0.953	0.108	-0.470	0.670	0.762	1.19
Median income (natural log)	0.901	0.0308	-3.05	0.002	0.843	1.0
Age (months; reference category 0-8 months)						
9-14	1.14	0.0870	1.69	0.091	0.980	1.32
15-19	1.41	0.112	4.27	<0.001	1.20	1.64
20-29	1.36	0.105	3.94	<0.001	1.17	1.58
30+	1.35	0.109	3.75	<0.001	1.15	1.58
Year home was built (reference period: before 1950)						
1950-1977	0.780	0.0333	-5.85	<0.001	0.717	0.848
1978-1987	0.691	0.0303	-8.38	<0.001	0.634	0.753
1988-1997	0.658	0.0289	-9.52	<0.001	0.604	0.717
1998-2002	0.618	0.0278	-10.7	<0.001	0.566	0.675
2003-2007	0.653	0.0298	-9.36	<0.001	0.597	0.714
2008 and later	0.619	0.0361	-8.24	<0.001	0.552	0.694
Season (reference: winter)						
Spring	0.985	0.0173	-0.880	0.380	0.951	1.02
Summer	1.03	0.0178	1.96	0.050	1.00	1.07
Fall	0.976	0.0186	-1.27	0.204	0.940	1.01

Table S2.

Coefficients from mixed-effects Tobit regression model of log-transformed children's blood lead levels in Wake County NC (left-censored at log[1 ug/dL]). Coefficients are for a model including all variables shown (*n*=45,045 complete observations). The model accounts clustering within households and census block groups.

Independent Variable	Exponentiated Coefficient (Odds Ratio)	Std. Err.	*Z*	*p>Z*	Lower 95% CI	Upper 95% CI
Private well	1.25	0.105	2.68	0.007	1.06	1.48
Male biological sex	1.03	0.0687	0.450	0.654	0.904	1.17
Home value (natural log of total $ value)	0.619	0.0611	-4.85	<0.001	0.510	0.752
Born in extraterritorial jurisdiction (ETJ) (reference category: not born in ETJ)	1.22	0.146	1.68	0.092	0.968	1.55
Blood draw year – 2002 (natural log)	0.0109	0.00204	-24.1	<0.001	0.00756	0.0157
Home area (sq. ft.)	1.00	4.75E-05	-0.380	0.703	1.00	1.00
Blood draw type (venous=0; capillary=1)	0.875	0.0881	-1.32	0.186	0.719	1.07
African American population proportion	0.905	0.171	-0.530	0.596	0.624	1.31
Hispanic population proportion	0.375	0.191	-1.92	0.05	0.138	1.02
Median income (natural log)	0.757	0.0929	-2.27	0.023	0.595	0.963
Age (months; reference category 0-8 months)						
9-14	0.658	0.206	-1.34	0.180	0.357	1.21
15-19	1.15	0.373	0.420	0.672	0.607	2.17
20-29	0.927	0.306	-0.230	0.818	0.486	1.77
30+	0.985	0.331	-0.0400	0.965	0.510	1.90
Year home was built (reference year: 1950)						
1950-1977	0.433	0.0578	-6.27	<0.001	0.333	0.563
1978-1987	0.421	0.0600	-6.07	<0.001	0.319	0.557
1988-1997	0.355	0.0513	-7.16	<0.001	0.268	0.472
1998-2002	0.329	0.0497	-7.36	<0.001	0.245	0.442
2003-2007	0.327	0.0522	-7.00	<0.001	0.239	0.447
2008 and later	0.650	0.152	-1.85	0.065	0.412	1.03
Season (reference: winter)						
Spring	1.00	0.0881	0.0100	0.991	0.842	1.19
Summer	1.18	0.102	1.96	0.051	1.00	1.40
Fall	1.10	0.107	0.940	0.345	0.905	1.33

Table S3.

Coefficients from mixed-effects logit regression estimating whether children's blood lead levels equal or exceed the CDC action level of 5 µg/dL in Wake County, NC. Coefficients are for a model including all variables shown (*n*=45,266 complete observations). The model accounts for clustering within households and census block groups.

Independent Variable	Exponentiated Coefficient (Odds Ratio)	Std. Err.	*Z*	*p>Z*	Lower 95% CI	Upper 95% CI
Private well	1.47	0.0721	7.95	<0.001	1.34	1.62
Male biological sex	1.10	0.0276	3.83	<0.001	1.05	1.16
Home value (natural log of total $ value)	0.574	0.0240	-13.3	<0.001	0.529	0.623
Born in extraterritorial jurisdiction (ETJ) (reference category: not born in ETJ)	1.18	0.0829	2.34	0.019	1.03	1.35
Blood draw year – 2002 (natural log)	0.00388	0.000415	-51.89	<0.001	0.00314	0.00478
Home area (sq. ft.)	1.00	4.64E-06	5.17	<0.001	1.00	1.00
Blood draw type (venous=0; capillary=1)	1.32	0.0706	5.19	<0.001	1.19	1.47
African American population proportion	1.74	0.191	5.05	<0.001	1.40	2.16
Hispanic population proportion	1.02	0.256	0.0800	0.934	0.625	1.67
Median income (natural log)	0.808	0.0590	-2.92	0.004	0.700	0.932
Age (months; reference category 0-8 months)						
9-14	1.44	0.224	2.33	0.020	1.06	1.95
15-19	2.20	0.361	4.79	<0.001	1.59	3.03
20-29	2.08	0.331	4.58	<0.001	1.52	2.84
30+	2.02	0.338	4.19	<0.001	1.45	2.80
Year home was built (reference year: 1950)						
1950-1977	0.623	0.0574	-5.14	<0.001	0.520	0.746
1978-1987	0.478	0.0450	-7.83	<0.001	0.398	0.575
1988-1997	0.430	0.0410	-8.85	<0.001	0.357	0.519
1998-2002	0.369	0.0360	-10.2	<0.001	0.305	0.447
2003-2007	0.426	0.0412	-8.82	<0.001	0.352	0.515
2008 and later	0.379	0.0450	-8.17	<0.001	0.301	0.479
Season (reference: winter)						
Spring	0.952	0.0364	-1.28	0.199	0.884	1.03
Summer	1.06	0.0398	1.62	0.105	0.987	1.14
Fall	0.934	0.0384	-1.65	0.098	0.862	1.01

Table S4.

Coefficients from mixed-effects logit regression estimating whether children's blood lead levels exceed the population median (1 µg/dL) in Wake County, NC. Coefficients are for a model including all variables shown (*n*=45,266 complete observations). The model accounts for clustering within households and census block groups.

Variable	Children in Homes with Private Wells (n=7,709)	Children in Homes Served by a Regulated Water Utility (n=43,982)
Age		
< 9 months	0.78%	0.85%
9-14 months	70.55%	70.59%
15-19 months	6.95%	6.50%
20-29 months	14.23%	14.30%
≥30 months	7.49%	7.77%
Biological sex		
Female	49.52%	49.24%
Male	50.48%	50.76%
Year tested***		
2002-2006	20.98%	18.55%
2007-2012	42.78%	44.79%
2013-2017	36.24%	36.66%
Year house was built***		
Before 1950	4.80%	5.95%
1950-1978	23.29%	20.80%
1978-1988	16.05%	14.74%
1988-1998	18.07%	15.87%
1998-2003	13.53%	18.04%
2003-2008	15.26%	18.54%
2008 or later	9.00%	6.06%
Home value***		
Quartile 1 (<$132,880)	31.82%	23.78%
Quartile 2 ($132,881-188,281)	23.69%	25.68%
Quartile 3 ($188,282-292,135)	20.62%	25.88%
Quartile 4 (>$292,135)	23.87%	24.65%
Black population proportion***		
Quartile 1 (<0.095)	28.09%	25.16%
Quartile 2 (.095-0.211)	28.12%	24.42%
Quartile 3 (0.212-0.370)	27.10%	24.29%
Quartile 4 (>0.370)	16.69%	26.12%
Median income***		
Quartile 1 (<$49,012)	15.37%	26.39%
Quartile 2 ($49,012-62,925)	27.03%	24.88%
Quartile 3 ($62,926-83,985)	33.70%	23.47%
Quartile 4 (>$83,985)	23.90%	25.27%

Variable	Children in Homes with Private Wells (n=7,709)	Children in Homes Served by a Regulated Water Utility (n=43,982)
Blood draw type		
Capillary	9.10%	9.66%
Venous	90.90%	90.34%
Season		
Winter	22.92%	23.05%
Spring	25.48%	25.22%
Summer	25.85%	25.84%
Fall	25.75%	25.90%
***Distributions between children with private wells and those with city water differs significantly, $p<0.001$, as measured by chi-square tests.		

Table S5.

Distributions of covariates used in regressions by water source.

Stata Code

Code for Merging Blood Lead and Property Parcel Tax Data

```
*Concat two address strings
gen address_c=siteaddr
replace address_c=siteaddres if missing(siteaddr)

gen zipcode=.
replace zipcode=zipnum
gen cityr = citydecode

gen cityr2= cityr
replace cityr2 = "Angier" if cityr=="ANG" | cityr=="ANGIER"
replace cityr2 = "Apex" if cityr=="APE" | cityr=="APEX"
replace cityr2 = "Cary" if cityr=="CAR" | cityr=="CARY"
replace cityr2 = "Durham" if cityr=="DUR" | cityr=="DURHAM"
replace cityr2 = "Garner" if cityr=="GAR" | cityr=="GARNER"
replace cityr2 = "Holly Springs" if cityr=="HOL" | cityr=="HOLLY SPRINGS"
replace cityr2 = "Knightdale" if cityr=="KNI" | cityr=="KNIGHTDALE" |
cityr=="KKNIGHTDALE" | cityr=="KNIGHTDAL" | cityr=="KNIGHTDALEE"
replace cityr2 = "Morrisville" if cityr=="MOR" | cityr=="MORRISVILLE" |
cityr=="ORRISVILLE" | cityr=="MARRISVILLE" | cityr=="MIRRISVILLE" |
cityr=="MOORESVILLE" | cityr=="MOORISVILLE" | cityr=="MOORRISVILLE" |
cityr=="MORISSVILLE" | cityr=="MORISVILLE" | cityr=="MORRISIVLLE" |
cityr=="MORRISVILLEA" | cityr=="MORRISVLLE" | cityr=="MORRIVILLE" |
cityr=="MORRRISVILLE" | cityr=="MORRSIVILLE" | cityr=="MORRSVILLE"
replace cityr2 = "Raleigh" if cityr=="RAL" | cityr=="RALEIGH" | cityr=="RAELIGH" |
cityr=="RALEIGHJ" | cityr=="RALEIGHT" | cityr=="RALIEGH" | cityr=="RALIGH" |
cityr=="ROAEIGH" | cityr=="WALEIGH"
replace cityr2 = "Rolesville" if cityr=="ROL" | cityr=="ROLESVILLE" |
cityr=="ROLESVILLES"
replace cityr2 = "Wake Forest" if cityr=="WAK" | cityr=="WAKE FOREST" |
cityr=="WAKWE FOREST"
replace cityr2 = "Wendell" if cityr=="WEN" | cityr=="WENDELL"
replace cityr2 = "Willow Springs" if cityr=="WILLOW SPRING" |
cityr=="WILLOWSPRINGS" | cityr=="WILLOW SPRINGS" | cityr=="Willowsprings" |
cityr=="Willow Spring"
replace cityr2 = "Zebulon" if cityr=="ZEB" | cityr=="ZEBULON"
replace cityr2 = "Fuquay-Varina" if cityr=="FUQUAY VARINA" |
cityr=="FUQUAYVARINA" | cityr=="FUQ" | cityr=="FUQUAY VARINA" | cityr=="Fuquay
Varina" | cityr=="FUQUAY-VARINA"
replace cityr2 = "Clayton" if cityr=="CLAYTON" | cityr=="CLAY" | cityr=="Clay"
tab cityr2

nysiis cityr2, generate(cityr2_nysiis)
```

```
*Drop the missing addresses
drop if missing(address_c)
gsort address_c township -year

*Drop the duplicates addresses that have the same township
by address_c township, sort: gen seq=_n
keep if seq==1

capture drop dup
duplicates t address_c , gen(dup)
tab dup
order address_c

*Deleted potentially valid addresses with duplicate street, but different township
by address_c, sort: keep if _n==_N
gen idfm=_n

capture drop multiunit
gen multiunit=(land_class2018=="APT" | land_class2018=="COX" | totvalassd>10000000 |
structure>1199)
replace multiunit=. if missing(land_class2018) | totvalassd==. | structure==.

*Create a first word of each address
capture drop address_part*
gen address_part1=substr(address_c,1,strpos(address_c," "))
gen address_part3=substr(address_c,strrpos(address_c," "),.)
gen address_part2=substr(address_c,strpos(address_c," "),length(address_c)-
length(address_part1)-length(address_part3)+1)

gen address_part3_r = address_part3
replace address_part3_r = "CIRCLE" if address_part3=="CIR"
replace address_part3_r = "CROSSING" if address_part3=="CROSS" | address_part3=="CRSG"
replace address_part3_r = "DRIVE" if address_part3=="DR"
replace address_part3_r = "LANE" if address_part3=="LN"
replace address_part3_r = "STREET" if address_part3=="ST"
replace address_part3_r = "ROAD" if address_part3=="RD"
replace address_part3_r = "TERRACE" if address_part3=="TER"
replace address_part3_r = "GREEN" if address_part3=="GRN"
tab address_part3_r

nysiis address_part2, generate(address_part2_nysiis)

*Gen address parcel for comparison
gen address_c_parcel = address_c

address_part2_nysiis cityr_nysiis
```

```
drop if missing(address_c)
save "C:\Users\fishermb\Desktop\temp\final_merged_c2_v4", replace

clear

*Next steps involved importing county tax parcel data

*reshape wide date

gen address_c=" "
replace address_c = street1
drop if missing(address_c)

*first part of address
gen address_part1=substr(address_c,1,strpos(address_c," "))
*second part of address
gen address_part3=substr(address_c,strrpos(address_c," "),.)

*Middle part of address
gen address_part2=substr(address_c,strpos(address_c," "),length(address_c)-
length(address_part1)-length(address_part3)+1)

gen address_part3_r = address_part3
replace address_part3_r = "CIRCLE" if address_part3=="CIR"
replace address_part3_r = "CROSSING" if address_part3=="CROSS" | address_part3=="CRSG"
replace address_part3_r = "DRIVE" if address_part3=="DR"
replace address_part3_r = "LANE" if address_part3=="LN"
replace address_part3_r = "STREET" if address_part3=="ST"
replace address_part3_r = "ROAD" if address_part3=="RD"
replace address_part3_r = "TERRACE" if address_part3=="TER"
replace address_part3_r = "GREEN" if address_part3=="GRN"
tab address_part3_r

nysiis address_part2, generate(address_part2_nysiis)

gen cityr = city

gen cityr2= cityr
replace cityr2 = "Angier" if cityr=="ANG" | cityr=="ANGIER"
replace cityr2 = "Apex" if cityr=="APE" | cityr=="APEX"
replace cityr2 = "Cary" if cityr=="CAR" | cityr=="CARY"
replace cityr2 = "Durham" if cityr=="DUR" | cityr=="DURHAM"
replace cityr2 = "Garner" if cityr=="GAR" | cityr=="GARNER"
replace cityr2 = "Holly Springs" if cityr=="HOL" | cityr=="HOLLY SPRINGS"
replace cityr2 = "Knightdale" if cityr=="KNI" | cityr=="KNIGHTDALE" |
cityr=="KKNIGHTDALE" | cityr=="KNIGHTDAL" | cityr=="KNIGHTDALEE"
```

```
replace cityr2 = "Morrisville" if cityr=="MOR" | cityr=="MORRISVILLE" |
cityr=="ORRISVILLE" | cityr=="MARRISVILLE" | cityr=="MIRRISVILLE" |
cityr=="MOORESVILLE" | cityr=="MOORISVILLE" | cityr=="MOORRISVILLE" |
cityr=="MORISSVILLE" | cityr=="MORISVILLE" | cityr=="MORRISIVLLE" |
cityr=="MORRISVILLEA" | cityr=="MORRISVLLE" | cityr=="MORRIVILLE" |
cityr=="MORRRISVILLE" | cityr=="MORRSIVILLE" | cityr=="MORRSVILLE"
replace cityr2 = "Raleigh" if cityr=="RAL" | cityr=="RALEIGH" | cityr=="RAELIGH" |
cityr=="RALEIGHJ" | cityr=="RALEIGHT" | cityr=="RALIEGH" | cityr=="RALIGH" |
cityr=="ROAEIGH" | cityr=="WALEIGH"
replace cityr2 = "Rolesville" if cityr=="ROL" | cityr=="ROLESVILLE" |
cityr=="ROLESVILLES"
replace cityr2 = "Wake Forest" if cityr=="WAK" | cityr=="WAKE FOREST" |
cityr=="WAKWE FOREST"
replace cityr2 = "Wendell" if cityr=="WEN" | cityr=="WENDELL"
replace cityr2 = "Willow Springs" if cityr=="WILLOW SPRING" |
cityr=="WILLOWSPRINGS" | cityr=="WILLOW SPRINGS" | cityr=="Willowsprings" |
cityr=="Willow Spring"
replace cityr2 = "Zebulon" if cityr=="ZEB" | cityr=="ZEBULON"
replace cityr2 = "Fuquay-Varina" if cityr=="FUQUAY VARINA" |
cityr=="FUQUAYVARINA" | cityr=="FUQ" | cityr=="FUQUAY VARINA" | cityr=="Fuquay
Varina" | cityr=="FUQUAY-VARINA"
replace cityr2 = "Clayton" if cityr=="CLAYTON" | cityr=="CLAY" | cityr=="Clay"
tab cityr2

nysiis cityr2, generate(cityr2_nysiis)

gen zipcode = zip

*Gen address DPH for comparison
gen address_c_DPH = address_c

dtalink address_c 20 -1 cityr2 1 -5 cityr2_nysiis 1 -7 address_part1 2 -20 address_part2 10 -2
address_part3_r 5 -2  address_part2_nysiis 5 -5  zipcode 1 -5  ///
using "C:\Users\fishermb\Desktop\temp\final_merged_c2_v4.dta", cutoff(7) bestmatch describe
examples(25) block(zipcode)

drop if _matchflag!=1

ds _matchID, not
collapse (lastnm) `r(varlist)', by(_matchID)
```

Data Analysis Code

```
*Recode some of the variables
gen tapyear=test_year-2001
gen medhouseholdincome_r = medhouseholdincome/10000
gen home_value_r = home_value/100000
gen well_water =.
replace well_water = private_well

*Mixed-effects Tobit regression model predicting log(BLL)
metobit log_lead i.private_well gender_bin ///
 log_home_value born_in_etj log_test_year home_area_sqft blood_draw_type proportion_black
proportion_hispanic logincome ///
i.age_months_cat i.year_of_construction i.season || match_addr:, ll(0) vce(cluster block_group)

*Logit regression predicting odds of BLL>5 µg/dL
xtlogit actionlevel private_well gender_bin ///
 log_home_value born_in_etj log_test_year home_area_sqft blood_draw_type proportion_black
proportion_hispanic logincome ///
i.age_months_cat i.year_of_construction i.season, re i (match) vce(cluster block_group) or

*Logit regression predicting odds of BLL above population median
xtlogit bll_binary_2ugdl private_well gender_bin ///
 log_home_value born_in_etj log_test_year home_area_sqft blood_draw_type proportion_black
proportion_hispanic logincome ///
i.age_months_cat i.year_of_construction i.season, re i (match) vce(cluster block_group) or
```