# Appendix A. Notation

* Census tracts are denoted by $i=1,...,I$
* Observations are denoted by $j=1,...,J$ for the $J$
* $y\_{j}$ - outcome ($e.g.$ smoking) for observation $j$
* $u\_{j}$ - is the census tract for observation $j$
* $w\_{j}$ - is the raked sampling weight for observation $j$
* $z\_{j}$ - zip code for observation $j$
* $P\_{i}$ -True census tract level population prevalence

# Appendix B. Imputation of Census Tract

The Department of Housing and Urban Development (HUD) provides the estimated proportion of a zip code's total residential addresses, which fall within each census tract (CT). For example, if zip code $z$ overlaps with 5 census tracts, we would have $p\_{z1},p\_{z2},p\_{z3},p\_{z4},$ and $p\_{z5}$ where

$$p\_{z1}=\frac{number of residential addresses in zip z and CT1}{total number of residential addresses in zip z}$$

and $\sum\_{i=1}^{5}p\_{zi}=1$. We assume that the probability of being from CT $i$ given you are in zip code $z$ is $p\_{zi}$.

For observations with missing census tracts, we impute a census tract based on a multinomial distribution with the HUD probabilities ($p\_{zi}$). For a single imputation each observation $j$ is assigned to census $i$ creating a set of indices $s\_{i}$. The imputation is repeated $D$ times, resulting in $d=1,...,D$ complete observation data sets (no missing census tracts).

# Appendix C. Hierarchical Bayesian Model

For each complete data set $d$ our approach is to summarize the data in census tract $i$ via the asymptotic distribution of the estimator of $P\_{i}$, which we denote $ˆ\_{i}=\sum\_{j\in s\_{i}}^{}w\_{j}y\_{j}/\sum\_{j\in s\_{i}}^{}w\_{j}$, the Hajek estimator (1) of $P\_{i}$, with corresponding variance estimator var$(ˆ\_{i})$. In this way the design is acknowledged in both the estimator and the variance. We define the area-level data summary as $x\_{i}=log\left[ˆ\_{i}/(1-ˆ\_{i})\right]$ as the empirical logistic transform of $ˆ\_{i}$. This approach constrains the probability to lie in (0,1). The likelihood is then taken as the asymptotic distribution

$$x\_{i}|P\_{i}∼N\left(log\left[\frac{P\_{i}}{1-P\_{i}}\right],\frac{ˆ(ˆ\_{i})}{ˆ\_{i}^{2}(1-ˆ\_{i})^{2}}\right).$$

We employ three-stage models with the first stage given by $x\_{i}|P\_{i}$ above, which was shown to perform well in a small area estimation context and has been applied to annual zipcode-level BRFSS data (2). When the $ˆ\_{i}=0/1$ we have $ˆ(ˆ\_{i})=0$ which is problematic in the first stage of our model. In these cases we use the method of moments based on a beta-binomial model described in the supplementary materials of (3) to provide $ˆ\_{i}$ and $ˆ(ˆ\_{i})$ that are adjusted to be non-zero.

At the second stage of the model we introduce the spatial random effects terms, corresponding to the convolution model of (4), and denote the area-specific parameters as

$$η\_{i}=μ+θ\_{i}+ϕ\_{i}$$

where $μ$ is the overall risk level, $ϕ\_{i}|σ\_{ϕ}^{2}∼\_{iid}N(0,σ\_{ϕ}^{2})$ is an independent census tract random effect, and $θ\_{i}$ following an intrinsic conditional autoregressive (ICAR) model (5). The ICAR model is a non-parametric, stochastic smoothing model with

$$θ\_{i}|θ\_{k},k\in ne(i),σ\_{θ}^{2}∼N\left(\overline{θ}\_{i},\frac{σ\_{θ}^{2}}{n\_{i}}\right),$$

where $ne(i)$ indexes the set of neighbors of area $i$, $n\_{i}$ is the number of such neighbors and $\overline{θ}\_{i}=\frac{1}{n\_{i}}Σ\_{k\in ne(i)}^{}θ\_{k}$ is the mean of the neighbors.

For the third stage we use assign Gamma priors on the spatial conditional precision $σ\_{θ}^{-2}$ and the $iid$ precision parameter $σ\_{ϕ}^{-2}$. The rate and shape parameters, 0.5 and 0.008, respectively, were selected such that the 95% range is on the interpretable $σ\_{θ}$ and $σ\_{ϕ}$ scale of (0.056,4.04). We use an improper flat prior on $α$.

# Appendix D. Combining Estimates

Our goal is to describe the posterior distribution of $η\_{i}$ given the observed data. We assume $u\_{o}$ are the set of observed census tracts and $u\_{m}$ are the missing census tracts

$$p(η\_{i}|y,z,u\_{o})=∫p(η,u\_{m}|y,z,u\_{o})p(u\_{m}|y,z,u\_{o})du\_{m}≈\frac{1}{D}\sum\_{d=1}^{D}p(η|y,z,u\_{o},u\_{m}^{(d)})$$

where $u\_{m}^{(d)}∼p(u\_{m}|y,z,u\_{o})$. In our current implementation $u\_{j,m}^{(d)}∼p(u\_{j,m}|z\_{j})$ is the multinomial distribution based on HUD data, but in principle could involve other covariates.

Given our set of $D$ smoothed estimates of each $η\_{i}$ we find the mean posterior estimate

$$E(η\_{i}|y,z,u\_{o})≈\frac{1}{D}\sum\_{d=1}^{D}ˆ\_{i,d}=ˆ\_{i}$$

where $ˆ\_{i,d}=E(η\_{i}|y,z,u\_{o},u\_{m}^{(d)})$ is the estimated $η\_{i}$ from the $d$th complete data set.

Similarly we find a variance

$$V(η\_{i}|y,z,u\_{o})≈\frac{1}{D}\sum\_{d=1}^{D}V\_{i,d}+\frac{1}{D-1}\sum\_{d=1}^{D}\left(ˆ\_{i,d}-ˆ\_{i}\right)^{2}=ˆ\_{i}$$

where $V\_{i,d}=V(η\_{i}|y,z,u\_{o},u\_{m}^{(d)})$ (the posterior variance of $η\_{i}$ based on the $d$th complete dataset). This variance estimate has contributions from within and between the sets of estimates (1).

Finally, estimates for census tract $i$ are derived from expit($ˆ$) and 95% credible intervals are generated using

$$\left[expit\left(ˆ\_{i}-1.96×\sqrt{ˆ\_{i}}\right),expit\left(ˆ\_{i}+1.96×\sqrt{ˆ\_{i}}\right)\right].$$

# Appendix E. Scatter Plots Comparing Direct Estimates with Smoothed Estimates By Sample Size

| Point Estimates | CI Width |
| --- | --- |
| Direct to Complete Case SmoothedC:\Users\songl\Dropbox\PHSKC_SAE\Results\CompareHT_Smooth.jpeg | Direct to Complete Case SmoothedC:\Users\songl\Dropbox\PHSKC_SAE\Results\IntervalWidth.jpeg |
| Direct to MI SmoothedC:\Users\songl\Dropbox\PHSKC_SAE\Results\CompareHT_MI_Smooth.jpeg | Direct to MI SmoothedC:\Users\songl\Dropbox\PHSKC_SAE\Results\IntervalWidth_Direct_MI_Smooth.jpeg |

**REFERENCES**

1. Hajek, J. (1971). Comments on a paper by D. Basu. In: *Foundations of Statistical*
2. *Inference* (eds: V.P. Godambe and D.A. Sprott). Toronto: Holt, Rinehart and Winston.
3. Mercer LD, Wakefield J, Pantazis A, Lutambi AA, Masanja H, and Clark SJ. “Space-Time Smoothing of Complex Survey Data: Small Area Estimation for Child Mortality”. *Annals of Applied Statistics.* 9(4) *(2015*): 1889-1905.
4. Chen, Cici, Jon Wakefield, and Thomas Lumely. "The use of sampling weights in Bayesian hierarchical models for small area estimation." *Spatial and spatio-temporal epidemiology* 11 (2014): 33-43.
5. Besag, Julian, Jeremy York, and Annie Mollié. "Bayesian image restoration, with two applications in spatial statistics." *Annals of the institute of statistical mathematics* 43.1 (1991): 1-20.
6. Rue, Havard, and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC Press, 2005.