

Spatial Multilevel Modelling in the Galveston Bay Recovery Study Survey



Mary E. Thompson, Gang Meng, Joseph Sedransk, Qixuan Chen,
and Rebecca Anthopolos

Abstract The Galveston Bay Recovery Study conducted a longitudinal survey of residents of two counties in Texas in the aftermath of Hurricane Ike, which made landfall on September 13, 2008 and caused widespread damage. An important objective was to chart the extent of symptoms of Post-Traumatic Stress Disorder (PTSD) in the resident population over the following months. Wave 1 of the survey was conducted between November 17, 2008 and March 24, 2009. Waves 2 and 3 consisted of two month and one year follow-ups, respectively. With the use of a stratified, 3-stage sampling design, data were collected from 658 residents. The first stage of sampling within strata was the selection of clusters, or *area segments*. Our objective is to model the course of the repeated PTSD measures as a function of individual characteristics and area segment, and to examine the analytical and visual evidence for spatial correlation of the area segment effect. To incorporate design information, our multilevel analysis uses the composite likelihood approach of Rao et al. (*Survey Methodology*, 39, 263–282, 2013) and Yi et al. (*Statistica Sinica*, 26, 569–587, 2016). We compare this with a Bayesian multilevel analysis and discuss the estimability of the model when the cluster-level variation has spatial dependence.

M. E. Thompson (✉) · G. Meng

Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, ON, Canada
e-mail: methomps@uwaterloo.ca; gmeng@uwaterloo.ca

J. Sedransk

University of Maryland, College Park, MD, USA
e-mail: jxs123@case.edu

Q. Chen

Department of Biostatistics, Columbia University, New York, NY, USA
e-mail: qc2138@cumc.columbia.edu

R. Anthopolos

New York University, New York, NY, USA
e-mail: Rebecca.Anthopolos@nyulangone.org

Keywords Bayesian analysis · Complex survey design · Longitudinal data · Multilevel model · Pairwise composite likelihood

1 Introduction

The Galveston Bay Recovery Study (GBRS) survey was conducted to study the impact of Hurricane Ike, which had made landfall at Galveston Bay on September 13, 2008. The survey took place in Chambers County and Galveston County in Texas. Galveston County includes Galveston Island and the Bolivar Peninsula, with Goat Island just to the north. The hurricane caused severe damage, particularly on the Bolivar Peninsula and Goat Island, but also on Galveston Island and further into the Bay.

With the intention of gathering data close to the time of the disaster, the investigators were able to design a three-wave longitudinal survey of which Wave 1 went into the field about two months after Hurricane Ike. Wave 1 continued until March 24, 2009. Wave 2 was a half-hour follow-up intended to be conducted two to three months after the initial interview. Wave 3 was a full follow-up survey intended to be conducted about a year after the first interview (University of Michigan Survey Research Center/Institute for Social Research 2010). The sampling design was a two-stage area sample of households from address-based frames, while interviewing took place by telephone. The main goal was to characterize trajectories and determinants of post-disaster mental health outcomes, such as Post-Traumatic Stress Disorder (PTSD), as measured through a severity score computed from responses to a 17-item scale (Pietrzak et al. 2013). Another aim (Gruebner et al. 2016a,b) was to use spatial analysis to identify patterns of mental health and wellness, and their predictors, across the geographic area in the aftermath of the disaster.

With a view to incorporating the complex features of the sampling design, Anthopolos et al. (2020) have proposed a Bayesian growth mixture model, where the three-wave trajectory of the log of the PTSD severity score is modelled within latent classes. The modelling of latent class membership is multilevel because of the clustering of the sample, and incorporates spatial dependence across adjacent clusters. Sampling design variables such as household size and auxiliary information on the frames are incorporated as covariates. Inference concerning the cluster-level variance components of latent class membership is part of the purpose.

The aim of this paper is to implement a frequentist approach to incorporating complex sampling design features in a more basic repeated measures analysis of the same data, where inference concerning the cluster-level variance components for the log PTSD severity score itself is envisaged. The complex sampling design features are incorporated using pairwise likelihood using the approach of Rao et al. (2013) and Yi et al. (2016).

Section 2 will describe the sampling design in detail. Sections 3 and 4 will document the construction of survey weights and the derivation of the inclusion probabilities required for the illustrative analyses. Section 5 specifies the spatial

multilevel model under consideration. Sections 6 and 7 present a standard Bayesian analysis and the proposed frequentist analysis, respectively. Section 8 discusses the advantages and disadvantages of the two approaches, with reference to the ways in which they use the information in the sampling design and the weights.

2 Sampling Design

The following description is taken from Valliant et al. (2009) and University of Michigan Survey Research Center/Institute for Social Research (2010).

There were two sampling frames. One frame was the Experian Gold list for Galveston and Chambers Counties, purchased from the credit reporting agency Experian. This list had demographic information that could be used in an attempt to identify households and persons with higher probability of experiencing PTSD in the short or long run, based on earlier studies. A score was then constructed by the SRC to classify most of the households as high risk or low risk (or with insufficient data to determine) for PTSD after a disaster. The other frame was an area probability frame created by field staff listing procedures. Its coverage was more comprehensive, for example, including growth since the 2000 Census.

For the GBRS survey, FEMA maps of the flooding in the Galveston area immediately after Hurricane Ike and Census 2000 data were used to divide the two-county area into five geographic strata:

Stratum 1: Galveston Island and the Bolivar Peninsula, which suffered storm surge damage

Stratum 2: Flooded areas of the mainland

Stratum 3: Non-flooded areas of the mainland which had relatively high rates of poverty in the 2000 Census

Stratum 4: Non-flooded, non-poverty areas east of Route 146 (and thus close to the Bay)

Stratum 5: Non-flooded, non-poverty areas west of Route 146 and the remainder of Chambers County (not flooded for the most part)

Within strata, the researchers constructed *area segments* composed of census blocks from the 2000 Census. Eighty (80) of these were to be selected. It was initially decided that the relative sampling rates in the strata would be 4, 4, 2, 2, and 1, so that Stratum 1 and Stratum 2 would be oversampled, while Stratum 5 would be undersampled. Implementing these rates resulted in an allocation of area segments to strata of 42, 4, 16, 4, and 14. Within strata, the area segments were selected with probability proportional to a size measure, namely the number of occupied housing units in the 2000 census.

Three of the selected segments in Stratum 1 were in an area (the Bolivar Peninsula) that received extensive damage and could not be field-listed. Thus the final numbers of segments represented in the strata samples are 39, 4, 16, 4 and 14; or 77 in all. Figure 1 shows the locations of the census tracts of the sampled area segments, coloured according to stratum, superimposed on a map of population

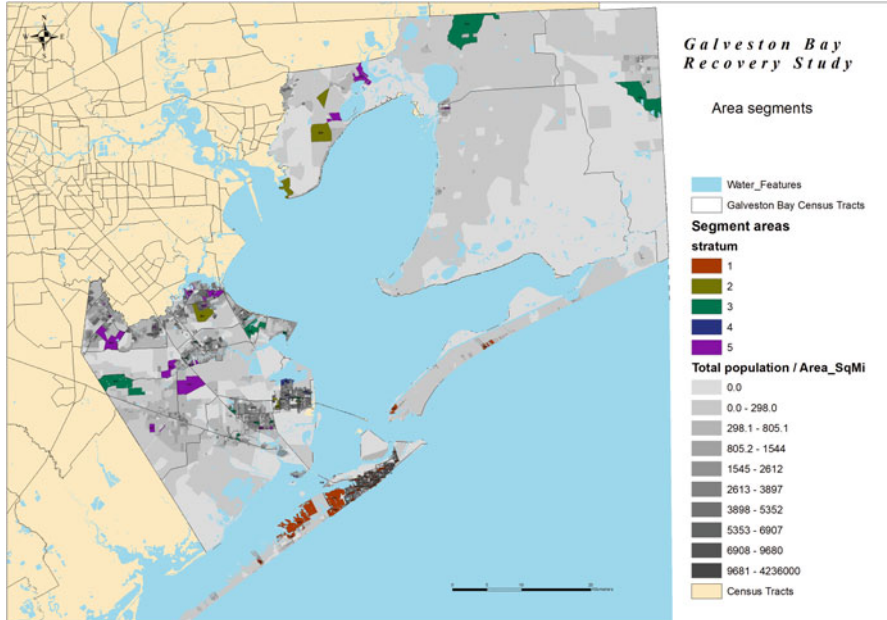


Fig. 1 Census blocks of the sampled area segments superimposed on a map of population density from the 2000 Census

density from the 2000 Census. From this map it is apparent that the sample is taken from areas of higher population density.

The area field listing included many housing units present in the area which did not appear on the Experian frame, and there were many cases where the same housing unit was recorded differently on the two frames. Within selected area segments, it was decided to use the Experian list as the primary sampling frame. Households therein were subdivided in each geographic stratum into *High Risk for PTSD* and *Other* (low risk or not determinable). The High Risk group was sampled at a rate 1.5 times that of the Other group. A separate sample was then taken from the subset of the area field frame listings which did not appear on the Experian frame. Whether this was the case could not always be determined perfectly: in some apartment buildings, there were cases that had a chance of selection on both frames. In the end, there were 124 Wave 1 interviews that came from the area frame (all coded as *other* for the risk variable) and 534 from the Experian frame.

In a *first phase* of sampling, selected households where it was possible to make contact were rostered, and in each, a member was selected at random from among those who were 18 years of age or older at the time of selection. *Respondent locating* was a major part of the effort, and this task was sent first to an outside vendor for internet locating of respondents, to be followed by in-person tracking.

In a *second phase* of sampling of households not responding in the first fieldwork period, cases from the first released sample, either in tracking or never contacted,

were considered for further effort aimed at completing an interview. Of 489 eligible cases, 250 were selected.

Overall there were 2116 selected housing units, 420 of which were determined to be out of scope, and 658 of which resulted in a completed interview. Twenty (20) of the selected respondents were judged ineligible. Thus the Wave 1 response rate was approximately 40%. Weighted re-interview rates were 81.4% at Wave 2 and 73.3% at Wave 3.

3 Survey Weights

Survey weights were constructed for the GBRS survey data. Only the Wave 1 weights will be described here. The process is described in Valliant et al. (2009). The initial household weight was calculated as the reciprocal of the intended household inclusion probability, taking into account risk status (High Risk for PTSD or Other), the possibility of inclusion in both frames, and phase of sampling.

Consideration of phase of sampling leads to high variability of the initial household weight within strata.

The household weight was then adjusted for non-response, as follows. Contact, screening, and main interview completion were modelled in terms of housing unit characteristics: observed damage to the unit, observed destruction of the unit, stratum, Bolivar indicator, Experian indicator, High Damage Area indicator, Median Year Housing Units Built (an area segment variable), Ever a Refusal (15% of household refusals were converted), and Number of Calls.

Four adjustment strata were created, and weights were adjusted by the mean predicted contact, screening, and interview propensities in their adjustment strata. A few non-response adjustment factors were very large, and the corresponding weights were trimmed, with the reduction in weights being distributed across the other cases.

For each individual respondent, the person-level weight was the product of the non-response-adjusted household weight and the number of adults aged 18 or over in the household.

4 Inferring Inclusion Probabilities from the Weights

In the data file the household weights and person weights were provided, giving us the unconditional inclusion probability for each household and for each individual. For an illustrative design-based multilevel analysis, we needed to assign an estimated inclusion probability to each sampled area segment, and an estimated inclusion probability to each sampled individual, conditional on their area segment being sampled. The sampling of area segments was done using probability proportional to size sampling, where size was the 2000 census number of occupied housing units in the area segment. We were able to obtain an approximate value

of the size of an area segment by summing the year 2000 occupied housing unit numbers of the census blocks of sampled households within the area segment, and adjusting the sum upward so that the totals over area segments in the strata would match known numbers. This produced an estimated size variable \hat{N}_{hj} for each Stratum h and area segment j . It should be noted that the designers of the sampling plan would have had access to the true size values.

If we denote the initial household weight for household k in area segment j and Stratum h by w_{hjk} , we can write the reciprocal of w_{hjk} as

$$\pi_{k|h} \pi_{j|h}, \quad (1)$$

where $\pi_{j|h}$ is the needed area segment inclusion probability and $\pi_{k|h}$ is the design inclusion probability of household k within sampled area segment j in Stratum h . The value of $\pi_{k|h}$ depends on the risk stratum (High Risk for PTSD vs Other) of the household.

The High Risk for PTSD vs Other variable is not included on the data set. However, within many area segments, the lower household weights follow a pattern: the lowest weights are about 2/3 of the next lowest weights. Thus it appears that the lowest weights may correspond to deliberate oversampling, and we have assumed that they belong to households that were sampled at a rate of 1.5 times the “usual” rate in the area segment. We have also noted that within strata, the inclusion probabilities for households from the area frame were a fixed multiple of the inclusion probability of lower risk households from the Experian frame. Using these facts, together with information about the inclusion probabilities for the second phase samples, and additional assumptions, we have assigned a value of the High Risk indicator to each household.

Except for some extreme values due to the second phase of sampling, the initial household weights are not highly variable within area segments, and we approximated household inclusion probabilities by assuming simple random sampling within risk indicator value to begin with. Let a_{hj} (to be estimated) be N_{hj} times the sampling rate for lower risk households in area segment j of stratum h , where N_{hj} is the number of census 2000 occupied housing units in the area segment (the “size” of the area segment), so that for household k , the probability of inclusion $\pi_{k|h}$ is $1.5^{\delta_{hjk}} a_{hj} / N_{hj}$ where $\delta_{hjk} = 1$ if household k is of High Risk for PTSD, and 0 otherwise. Suppose the number of sampled households in the area segment is n_{hj} . Let the proportion of those households that appear (from their weights) to be High Risk for PTSD be \hat{p}_{hj} . Then a_{hj} can be estimated from the equation

$$n_{hj} = \hat{p}_{hj}(1.5\hat{a}_{hj}) + (1 - \hat{p}_{hj})\hat{a}_{hj}. \quad (2)$$

This gives a preliminary estimate of $N_{hj}\pi_{k|h}$ for each household k in the sample in area segment j .

Taking this estimate and multiplying by initial household weight, i.e., the reciprocal of the expression in (1), we computed a household-specific preliminary estimate of $\pi_{j|h}/N_{hj}$. We averaged these over the households with non-extreme

weights in area segment j to estimate $\pi_{j|h}/N_{hj}$. We multiplied by \hat{N}_{hj} and took the minimum of the result and 1 to obtain an approximate value of $\pi_{j|h}$ for each area segment j .

We then estimated the post-nonresponse inclusion probability for a household, given inclusion of its area segment, as the reciprocal of (the non-response adjusted household weight times the approximate value of the area segment inclusion probability). If we set aside three area segments with unusually high inclusion probabilities in Stratum 1, the average estimated area segment inclusion probabilities in the five strata are, respectively, 0.516, 0.689, 0.239, 0.260, and 0.113. The relative values of these are not very different from those of the initial target sampling rates, which were to be proportional to 4, 4, 2, 2, and 1.

These calculations allowed us to construct, for illustrative purposes, approximate decompositions of the person-level inclusion probabilities as follows:

$$\pi_{hjk} = \pi_{j|h} \pi_{k|hj} \pi_{i|hjk},$$

where $\pi_{i|hjk}$ is the reciprocal of the number of people aged 18 or over in household k . In what follows it will be convenient most of the time so suppress the stratum index h and combine the selection of household and person, writing the inclusion probability of person i of area segment j as

$$\pi_{ji} = \pi_j \pi_{i|j}. \tag{3}$$

5 Spatial Multilevel Model

Let the outcome variable y_{jit} be the logarithm of self-reported Post-Traumatic Stress Disorder (PTSD) severity score for resident i living in sampling cluster j at Wave t , $t = 1, 2, 3$. The sampling clusters are taken to be the area segments. We suppress notation for the sampling stratum h and the household k for simplicity. The PTSD severity scores were calculated as the sum of responses to 17 symptoms of PTSD, such as “repeated, disturbing memories of Hurricane Ike,” using the Checklist-Specific version (PCL-S) (Blanchard et al. 1996) with each symptom rated from 1 (not at all) to 5 (extremely). Questions were asked in reference to the period since the hurricane at Wave 1, and the period since the previous interview at Waves 2 and 3. Let \mathbf{x}_{ji} be the row vector of p covariates of interest for resident i in cluster j , potentially including age, gender, ethnicity, highest education completed, pre-disaster trauma exposure, pre-disaster PTSD, pre-disaster depression, hurricane-related trauma and stressors, peri-event reactions, and community-level social assets (Gruebner et al. 2016a). The model for the outcome variable could also depend on the sampling design through the sampling stratum, and through determinants of w_{ij} , such as a smooth function of the logarithm of the size variable (the number of occupied housing units in the sampled census blocks of the area segment); the risk

indicator for the household; the number of adult members of the household; and a function of the household non-response adjustment (Anthopolos et al. 2020). We have used all of these except the function of the size variable, this being omitted to keep the covariate space relatively simple.

The goal of this modelling approach is to examine risk factors, analytically and visually, associated with post-disaster scores of PTSD after accounting for longitudinal dependence, spatial correlation and the complex survey design. We propose a three-level model, where the three levels are the spatial cluster (the area segment), the individual within a cluster, and the survey wave within an individual. By an extension of notation j is the identifier of the *adjusted census tract* (CT) containing cluster j . Adjusted CTs were defined as follows: if two or more area segments (clusters) were in one official CT, the CT was split based on the number of area segments within it so that each area segment is in just one adjusted CT; if a CT has no area segment within it, that CT is combined with the nearest adjusted CT; thus after adjustment, the whole study area has the same number of adjusted CTs as the number of area segments.

The model for the outcome variable can be written as follows:

$$y_{jit} | \mu_{jit} \sim N(\mu_{jit}, \sigma_c^2)$$

$$\mu_{jit} = \alpha_{0ji} + \gamma_{02}I(t = 2) + \gamma_{03}I(t = 3)$$

$$\alpha_{0ji} = \beta_{0j} + \mathbf{x}_{ji}\beta + v_{0ji}$$

$$v_{0ji} \sim N(0, \sigma_{v0}^2)$$

$$\beta_{0j} = \gamma_{00} + w_{0j} + u_{0j}$$

$$w_{0j} \sim N(0, \sigma_{w0}^2),$$

where μ_{jit} is the expected value for individual i in area segment j at Wave t , and σ_c^2 is the within person variance component; the individual level intercept α_{0ji} depends on the covariates, and has an individual level variance component σ_{v0}^2 ; its area segment level intercept β_{0j} has the sum of two error terms, a spatially correlated term u_{0j} and an i.i.d. error term w_{0j} with variance component equal to σ_{w0}^2 .

For the spatially correlated error term of the area segment level intercept a relatively simple choice is the intrinsic conditional autoregressive (ICAR) prior (Besag et al. 1991):

$$u_{0j}|u_{0j'}, j' \in \text{ne}(j) \sim N\left(\bar{u}_{0j}, \frac{\sigma_{u0}^2}{n_j}\right), \tag{4}$$

where $\text{ne}(j)$ is the set of adjusted CTs which are neighbours of Area j , n_j is the number of such neighbours, and \bar{u}_{0j} is the mean of the neighbouring spatial random effects.

In this spatial multilevel model, we model the spatial dependence of clusters by the neighbourhood structure. Adjusted CTs are considered to be *neighbours* if they have a boundary edge or a corner in common.

A Note on Identifiability

An important reason for application of a multilevel spatial model is to try to estimate the response variable cluster means and to map them. Separating the cluster random effects into spatially correlated and independent parts can also be of interest, and that means not only estimating the variance components σ_c^2 , σ_{v0}^2 , σ_{w0}^2 , and σ_{u0}^2 but also estimating (in a Bayesian analysis) or predicting (in a frequentist analysis) w_{0j} and u_{0j} . With the model of this section and the kind of data available from the Galveston Bay Recovery Study, the variance components are identifiable in a frequentist likelihood analysis, or in the composite likelihood approach of Sect. 7. The quantities β_{0j} and γ_{00} are also estimable if the β_{0j} are constrained to have mean value 0. However, the separation of the random effect $w_{0j} + u_{0j}$ into its two components is not identifiable in these contexts. (The Bayesian analysis of Sect. 6 would allow such a separation because of the prior assumption on the variance components.) See Eberly and Carlin (2000) and Best et al. (2005) for discussions of this non-identifiability of spatial and random effects.

Leroux et al. (1999) and MacNab (2003) proposed a different model for $b_{0j} = w_{0j} + u_{0j}$ under which this total cluster random effect can be estimated, as well as a parameter λ that expresses the extent of spatial dependence of the cluster means. In this model, the covariance of b_{0j} and $b_{0j'}$ is the jj' -th element of the matrix $[\sigma_{w0}^2 + \sigma_{u0}^2][\lambda(D - A) + (1 - \lambda)I]^{-1}$ where D is the diagonal matrix with j -th entry equal to n_j , the number of neighbours of area segment j ; A is the adjacency matrix for the area segment clusters; and I is the identity matrix. Our method in this paper could be adapted to working with this parameterization.

6 A Bayesian Analysis

If a Bayesian approach is taken, for example, using WINBUGS, the following prior distributions for the parameters may be adopted:

$$\gamma_{02} \sim N(0, 1000); \gamma_{03} \sim N(0, 1000); \beta \sim MVN(0, 1000 * I),$$

where I is the identity matrix with the same number of rows as the dimension of \mathbf{x}_{ij} and the component standard deviations σ_c , σ_{u0} , σ_{v0} and σ_{w0} have a *Cauchy*(25) distribution, where *Cauchy*(h) signifies a half-Cauchy distribution with scale parameter h . The parameter γ_{00} is given an improper uniform prior. All parameters are *a priori* independent. In all analyses in this paper, we assume dropout is not informative.

The results of the Bayesian analysis are displayed in Table 1. The level of PTSD is seen to decrease after Wave 1, and to rise a little between Wave 2 and Wave 3. The level of PTSD tends to be higher among females, and to increase with age; it is higher for minorities; higher for people with PTSD prior to Hurricane Ike; higher for people with Ike-related trauma; higher for people with peri-event

Table 1 Results of the Bayesian analysis

Beta values		Mean	SD	2.50%	97.50%
Intercept		3.1500	0.1695	2.8230	3.4890
Survey waves	Wave 2	-0.1107	0.0131	-0.1367	-0.0850
	Wave 3	-0.0835	0.0135	-0.1096	-0.0570
Sex	Male	-0.0408	0.0224	-0.0842	0.0032
Age groups	35-54	0.0215	0.0303	-0.0379	0.0799
	55 and over	0.1180	0.0335	0.0511	0.1832
Race	Black non-hisp	0.1226	0.0356	0.0529	0.1932
	Hispanic	0.0929	0.0317	0.0312	0.1557
	Other non-hisp	0.0328	0.0497	-0.0644	0.1314
Education	= high school	-0.1006	0.0353	-0.1690	-0.0313
	> high school	-0.1379	0.0327	-0.2017	-0.0741
# traumatic events prior to Ike	2-3	0.0179	0.0264	-0.0337	0.0701
	4+	0.0471	0.0295	-0.0103	0.1049
Depression prior to Ike	Yes	0.0359	0.0287	-0.0204	0.0929
PTSD prior to Ike	Yes	0.2017	0.0355	0.1328	0.2720
Ike-related trauma	Yes	0.1058	0.0342	0.0390	0.1734
Ike-related stress	Yes	0.0028	0.0499	-0.0951	0.0996
Peri-event emotional reactions	Medium	0.1608	0.0272	0.1087	0.2149
	High	0.3801	0.0285	0.3234	0.4359
Non-response adjustment		-0.0262	0.0625	-0.1503	0.0945
# adult household members	2-3	0.0371	0.0242	-0.0104	0.0851
	4+	0.0585	0.0629	-0.0657	0.1826
High PTSD risk indicator	Yes	0.0090	0.0232	-0.0361	0.0557
Average social support		0.0670	0.0541	-0.0397	0.1727
Average collective efficacy		-0.0635	0.0384	-0.1372	0.0132
Within person variance	σ_c^2	0.0474	0.0026	0.0427	0.0529
Within cluster variance	σ_{v0}^2	0.0456	0.0041	0.0381	0.0539
Between cluster variance (indep)	σ_{w0}^2	0.0033	0.0025	0.0001	0.0090
Between cluster variance (ICAR)	σ_{u0}^2	0.0064	0.0051	0.0007	0.0197

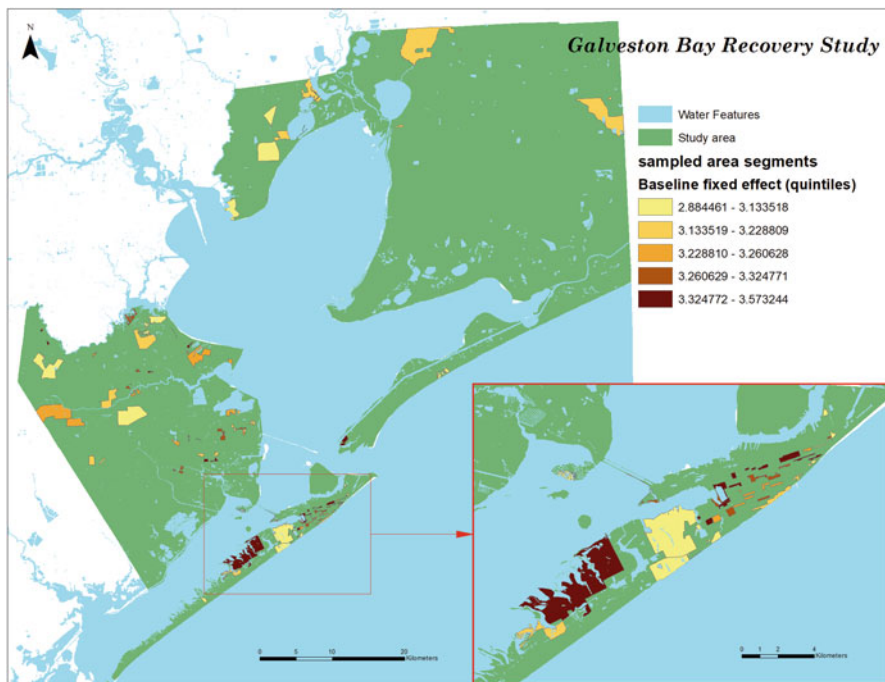


Fig. 2 Estimated cluster-level mean fixed effects of PTSD severity score

emotional reactions. The components of variance σ_c^2 and σ_{v0}^2 are estimated at 0.047 and 0.046, respectively, while the estimates of σ_{u0}^2 and σ_{w0}^2 are much smaller, and the posterior 2.50% quantiles of these last two variance components are very close to 0, suggesting that the variability within and between individuals dominates the area segment level variability.

Figures 2 and 3 display, respectively, the estimated cluster-level mean fixed effects and the estimated cluster-level random effects $u_0 + w_0$. (See Fig. 1 for comparison of the areas of high and low predicted PTSD severity with the stratum definitions.) The cluster-level mean fixed effects are the average, taken over sample members of the cluster at baseline, of the regression function with the coefficients replaced by their posterior mean values.

The estimated mean fixed effects have higher variability about their overall mean than do the estimated random effects. For the mean fixed effects, the values are mainly as expected given the characteristics of their strata. For example, higher values for the average predicted PTSD severity score are found in Stratum 1, in the eastern part of Galveston Island and Bolivar Island, and in some areas of Stratum 3, while lower values appear in parts of Stratum 5. The random effects also appear higher in Stratum 1.

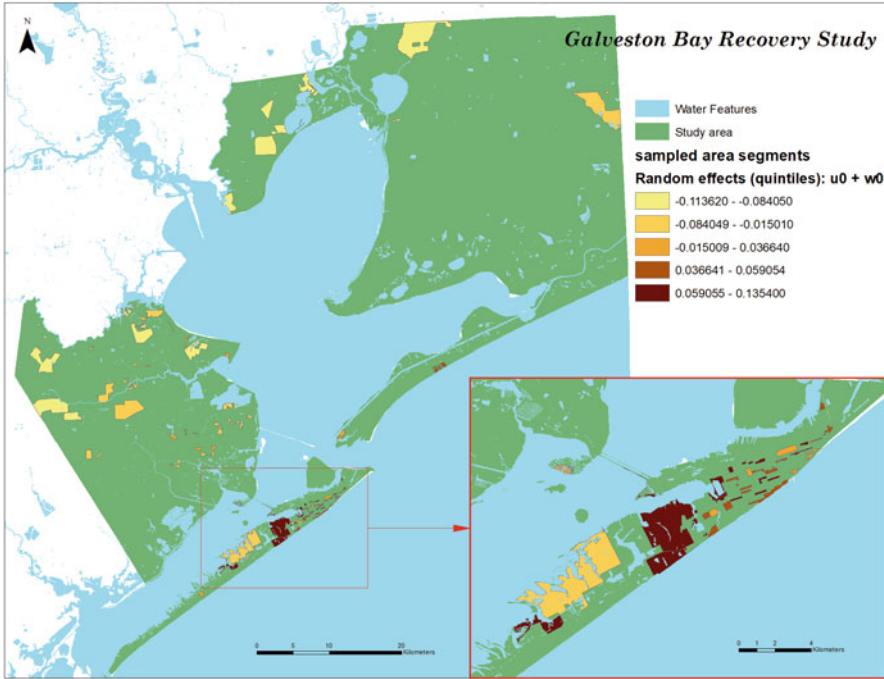


Fig. 3 Estimated cluster-level random effects of PTSD severity score

7 Frequentist Composite Likelihood Analysis

For a frequentist analysis, we consider adapting the weighted pairwise composite likelihood approach of Rao et al. (2013) and Yi et al. (2016). The idea in outline is as follows.

- Find (approximately) unbiased census estimating function terms for individual y values (for mean function parameters) and pairs of y values (for variance parameters).
- Combine them appropriately so that the combinations become maximum pairwise composite likelihood equations under the Gaussian model of Sect. 5.
- Estimate the census estimating functions by weighted sample estimating functions, and find their roots for point estimation.

7.1 Estimating Function System for Mean Parameters

For the mean parameters, the estimating function system could be a survey weighted GEE system:

$$\sum_j \sum_i w_{ji} X_{ji} (\Sigma_r)^{-1} (y_{ji} - M_{ji}), \tag{5}$$

where $M_{jit} = \gamma_{00} + \mathbf{x}_{ji} \beta + \gamma_{02} I(t = 2) + \gamma_{03} I(t = 3)$ is the marginal mean of y_{jit} , and $y_{ji} - M_{ji}$ is the vector of observed y_{jit} minus the corresponding M_{jit} ; X_{ji} is a $(p+3) \times 3$ matrix with columns equal to the transposes of $(\mathbf{x}_{ji}, 1, 1, 1)$, $(\mathbf{x}_{ji}, 0, 1, 0)$ and $(\mathbf{x}_{ji}, 0, 0, 1)$; and Σ_r is an exchangeable working correlation matrix, with 1's on the diagonal and with off-diagonal entries equal to a single correlation parameter ρ . The residuals are $\hat{z}_{jit} = y_{jit} - \hat{M}_{jit}$.

We note that the corresponding census estimating equation system is sub-optimal because the working covariance structure assumes independence of individuals, rather than the two-level model. Fitting this model using SUDAAN allows the stratification and two-stage design to be taken into account in estimation and testing hypotheses for the mean function parameters. This use of SUDAAN requires that the working correlation matrix be either exchangeable or independent.

7.2 Decomposition of the Error Term

The variance of

$$z_{jit} = y_{jit} - M_{jit} = e_{jit} + v_{0ij} + w_{0j} + u_{0j}$$

is

$$\sigma_c^2 + \sigma_{v0}^2 + \sigma_{w0}^2 + \sigma_{u_j}^2,$$

where $\sigma_{u_j}^2$ is the (unconditional) variance of u_{0j} under the ICAR model.

The covariance of z_{jit} and $z_{j't't'}$ is the (unconditional) covariance of u_{0j} , $u_{0j'}$ under the ICAR model. This is expressible as $C_{u_{jj'}}$, the jj' -th element of the matrix $\sigma_{u0}^2 (D - A)^{-1}$ (generalized inverse) where D is the diagonal matrix with j -th entry equal to n_j , the number of neighbours of area segment j ; and A is the adjacency matrix for the area segment clusters.

7.3 Estimating Equation System for Variance Components

If $z_{jit} = y_{jit} - M_{jit}$, and s_j denotes the sample of respondents in cluster j , the estimating equation system for the variance components can be written as:

$$\sum_j w_j \sum_{i \in s(j)} w_{i|j} \sum_{t=1}^3 (z_{jit}^2 - v_{1j}) = 0, \tag{6}$$

$$\sum_j w_j \sum_{i \in s(j)} w_{i|j} \sum_{t=1}^3 \sum_{t' < t} (z_{jit} z_{jit'} - v_{2j}) = 0, \tag{7}$$

$$\sum_j w_j \sum_{i' < i \in s(j)} w_{i'|j} \sum_{t=1}^3 \sum_{t'=1}^3 (z_{jit} z_{ji't'} - v_{3j}) = 0, \tag{8}$$

and

$$\sum_h \sum_{j' < j \in S_h} w_{jj'} \sum_{i' \in s(j')} \sum_{i \in s(j)} w_{i'|j'} w_{i|j} \sum_{t=1}^3 \sum_{t'=1}^3 (z_{jit} z_{ji't'} - C_{jj'}) = 0. \tag{9}$$

In Eqs. (6)–(8), the notation \sum_j signifies $\sum_h \sum_{j \in S_h}$, where S_h denotes Stratum h . In the system of equations (6)–(9)

$$v_{1j} = \sigma_{u_j}^2 + \sigma_{w_0}^2 + \sigma_{v_0}^2 + \sigma_c^2$$

$$v_{2j} = \sigma_{u_j}^2 + \sigma_{w_0}^2 + \sigma_{v_0}^2$$

$$v_{3j} = \sigma_{u_j}^2 + \sigma_{w_0}^2$$

$$C_{ujj'} = c_{jj'} \sigma_{u_0}^2$$

and $c_{jj'}$ is a known constant.

The solutions to (5) and to (6)–(9) have closed forms:

$$\hat{\sigma}_{u_0}^2 = \frac{\sum_h \sum_{j' < j \in S_h} w_{jj'} \sum_{i' \in s(j')} \sum_{i \in s(j)} w_{i'|j'} w_{i|j} \sum_{t=1}^3 \sum_{t'=1}^3 I_{it} I_{i't'} z_{jit} z_{ji't'}}{\sum_h \sum_{j' < j \in S_h} w_{jj'} \sum_{i' \in s(j')} \sum_{i \in s(j)} w_{i'|j'} w_{i|j} \ell_i \ell_{i'} c_{jj'}} \tag{10}$$

$$\hat{\sigma}_{w_0}^2 = \frac{\sum_j w_j \sum_{i \in s(j)} \sum_{i' < i \in s(j)} w_{i'|j} \sum_{t=1}^3 \sum_{t'=1}^3 I_{it} I_{i't'} (z_{jit} z_{ji't'} - a_j \hat{\sigma}_{u_0}^2)}{\sum_j w_j \sum_{i \in s(j)} \sum_{i' < i \in s(j)} w_{i'|j} \ell_i \ell_{i'}} \tag{11}$$

$$\hat{\sigma}_{w0}^2 = \frac{\sum_j w_j \sum_{i \in s(j)} w_{i|j} \sum_{t=1}^3 \sum_{t' < t} I_{it} I_{it'} (z_{jit} z_{jit'} - a_j \hat{\sigma}_{u0}^2)}{\sum_j w_j \sum_{i \in s(j)} w_{i|j} \ell_i (\ell_i - 1) / 2} - \hat{\sigma}_{w0}^2 \quad (12)$$

$$\hat{\sigma}_c^2 = \frac{\sum_j w_j \sum_{i \in s(j)} w_{i|j} \sum_{t=1}^3 I_{it} (z_{jit}^2 - a_j \hat{\sigma}_{u0}^2)}{\sum_j w_j \sum_{i \in s(j)} w_{i|j} \ell_i} - \hat{\sigma}_{w0}^2 - \hat{\sigma}_{v0}^2. \quad (13)$$

where I_{it} is the indicator function for i having an interview at t , and ℓ_i is the number of interviews of i .

7.4 Point Estimation

In a design-based analysis taking the weights to be the reciprocals of the corresponding inclusion probabilities, inclusion probabilities are needed for area segments j and for individuals within area segments $i | j$. Joint inclusion probabilities are needed for area segments jj' and for individuals within area segments $ii' | j$.

To illustrate the method with the GBRS data, having reconstructed inclusion probabilities from partial information on the data file as described in Section 4, we have used a Hájek approximation (Hájek et al. 1964) for the joint inclusion probabilities:

$$\pi_{ab} \simeq \pi_a \pi_b \left[1 - \frac{(1 - \pi_a)(1 - \pi_b)}{\sum_{a \in sample} (1 - \pi_a)} \right].$$

The paper by Haziza et al. (2008) gives an account of this and other joint inclusion approximations that can be used in variance estimation, including the one by Hartley and Rao (1962).

The residuals $z_{jit} = y_{jit} - M_{jit}$ were estimated using SUDAAN; the results of the SUDAAN analysis are displayed in Table 2:

In (5), the design weight for y_{ji} minus its marginal mean is the design weight for individual i in cluster j . This was taken to be the reciprocal of $\pi_{i|j} \pi_j$ as approximated in Sect. 4. These design weights were also applied in (13) and (12).

In (11), the design weight for $z_{ji't'} z_{jit}$ minus its marginal mean was taken to be the reciprocal of $\pi_j \pi_{ii'|j}$, where the second factor is the joint inclusion probability of i and i' , given that cluster j is included. The second factor was taken to be the product of the reciprocals of the numbers of adults in their households, times the joint inclusion probabilities of their households, given that cluster j is included. This last factor was calculated by a Hájek approximation from the individual conditional inclusion probabilities. Finally, in (10), the weight $w_{jj'}$ was taken to be the reciprocal of the Hájek approximation to the joint cluster inclusion probability $\pi_{jj'}$.

Table 2 Results of the GEE analysis

Beta values		Point est	SE	2.50%	97.50%
Intercept		3.0554	0.1995	2.6656	3.4452
Survey waves	Wave 2	-0.1165	0.0162	-0.1489	-0.0842
	Wave 3	-0.0785	0.0241	-0.1265	-0.0304
Sex	Male	0.0005	0.0262	-0.0518	0.0527
Age groups	35–54	0.0440	0.0283	-0.0123	0.1004
	55 and over	0.1277	0.0282	0.0715	0.1840
Race	Black non-hisp	0.2176	0.0326	0.1525	0.2827
	Hispanic	0.1324	0.0407	0.0513	0.2136
	Other non-hisp	0.0634	0.0457	-0.0277	0.1544
Education	= high school	-0.0097	0.0449	-0.0991	0.0797
	> high school	-0.0480	0.0443	-0.1363	0.0403
# traumatic events prior to Ike	2–3	-0.0353	0.0305	-0.0961	0.0254
	4+	-0.0065	0.0275	-0.0613	0.0484
Depression prior to Ike	Yes	0.0165	0.0295	-0.0424	0.0754
PTSD prior to Ike	Yes	0.1975	0.0311	0.1354	0.2595
Ike-related trauma	Yes	0.2115	0.0653	0.0814	0.3416
Ike-related stress	Yes	-0.0106	0.0470	-0.1043	0.0830
Peri-event emotional reactions	Medium	0.1458	0.0284	0.0891	0.2024
	High	0.3354	0.0372	0.2613	0.4095
Non-response adjustment		-0.0033	0.0597	-0.1223	0.1157
# adult household members	2–3	0.0449	0.0317	-0.0183	0.1082
	4+	0.0522	0.0620	-0.0715	0.1758
High PTSD risk indicator	Yes	0.0245	0.0270	-0.0293	0.0784
Average social support		0.1401	0.0355	0.0694	0.2108
Average collective efficacy		-0.1200	0.0436	-0.2069	-0.0332
Within person correlation	ρ	0.2903			

The point estimates of the first two variance components from the Galveston Bay data are $\hat{\sigma}_c^2 = 0.0381$ and $\hat{\sigma}_{w0}^2 = 0.0314$. The sum of $\hat{\sigma}_{u0}^2$ and $\hat{\sigma}_{w0}^2$, the total cluster-level variance component, is estimated at 0.0023, indicating that in this data set, the within cluster (between person) and within person variances dominate. We note that the sample has not been designed to facilitate a spatial analysis with the multilevel model of Sect. 5, and that the model itself may be too simple to apply well to the whole area.

The point estimates of variance components are somewhat smaller than those arising from the Bayesian analysis, but despite the high variability of the survey weights, the relative values from the frequentist analysis are similar.

There is good agreement between the estimates in Tables 1 and 2. Considering exclusion of zero from a nominal 95% interval as evidence of a non-zero effect, there are only four variables (average social support, average collective efficacy and the two education variables) where the inferences are different. It should be noted

that, although the GEE point estimation is sub-optimal, the standard errors from the GEE analysis do take into account the clustering and (through the design weights) the unequal probability sampling in the sampling design, and not surprisingly the SEs for the GEE analysis tend to be a little larger than the SDs of the Bayesian analysis.

7.5 Uncertainty Estimation

There are several methods that can be contemplated for the estimation of uncertainty in the point estimates arising from the system (5) and (6)–(8) or the system (5) and (10)–(13).

Applying a classical design-based approach would require the use of third and fourth order approximate inclusion probabilities. More appealing would be a model-based estimator of the mean-squared error of the design-based point estimators, using a *sandwich estimation* technique, described next in a simpler case.

Suppose $\hat{\theta}$ is the solution of the estimating equation

$$\sum_{i \in s} w_i \phi_i(y_i, \theta),$$

where under the model for the observations y_i , the terms $\phi_i(y_i, \theta)$ are correlated, with a correlation structure having parameters ρ . Then consider the Taylor series expression for the estimation error:

$$\hat{\theta} - \theta \simeq \left(\sum_{i \in s} w_i \frac{\partial \phi_i}{\partial \theta} \right)^{-1} \sum_{i \in s} w_i \phi_i.$$

The square of the left-hand side can be expressed as

$$(\hat{\theta} - \theta)^2 = \left(\sum_{i \in s} w_i \frac{\partial \phi_i}{\partial \theta} \right)^{-1} \sum_i \sum_{i'} w_i w_{i'} \phi_i \phi_{i'} \left(\sum_{i \in s} w_i \frac{\partial \phi_i}{\partial \theta} \right)^{-1}.$$

The factor in the middle can be replaced by its expectation in terms of the ρ parameters. The estimator of the variance of $\hat{\theta}$ could be the same expression evaluated at $\hat{\theta}$ and $\hat{\rho}$.

Another possible approach would be to treat the expressions in the left-hand sides for the sample-based maximum pairwise composite likelihood equations as analogous to the score function in a corresponding Gaussian model for the generation of the observations, as developed in the case of a simpler random effects model by Thompson et al. (2022). In that case, applying an adjustment to the curvature of the corresponding log likelihood along the lines of that proposed by

Ribatet et al. (2012) would make the pairwise log likelihood equations information unbiased and bring the inference based on them closer to that of a Bayesian analysis.

8 Discussion and Conclusions

We have outlined a frequentist design-based approach to estimation of the parameters of a multilevel repeated measures model with a continuous outcome, using data from a complex stratified three stage sampling design. The method uses the sample data to estimate population pairwise composite likelihood estimating functions. We have applied it to complex survey data from the Galveston Bay Recovery Study. In this application, the point estimates are broadly similar to those obtained from a Bayesian analysis of the same model.

Besides availability in software, important advantages of the Bayesian approach are the capacity to estimate the parameters of complex models and the principled expression of uncertainty through posterior distributions and credible intervals. From the frequentist perspective, a disadvantage of some Bayesian approaches is a requirement for knowledge of the variables influencing the sampling design, and the form of that influence. Incorporating this knowledge in the model accounts for the way in which the sampling design may distort the population relationships of interest. Other Bayesian approaches, such as the one used by Anthopolos et al. (2020), and used in Sect. 6, summarize the design features by including in the model the sample weights as covariates. When we include the survey design variables in the model, the interpretation of covariates of interest is altered, and may be changed in ways that do not align with scientific investigation.

The design-based frequentist approach attempts to address directly and compensate for this distortion. An advantage of this approach is that it can be applied in a straightforward manner to simple analytic uses of complex survey data with the use of a single set of survey weights supplied with the data. With this approach, there is also a natural extension to account for missing data by multiplying the baseline weight of someone who has responded at Wave t by the reciprocal of the probability of remaining in the sample up to that wave. Disadvantages are that the method as applied in this paper requires linear or quadratic estimating functions and that the variance components at the cluster-level tend to be weakly estimable.

We recommend the use of both methods for comparison in simple analytic uses of the data.

Acknowledgments The authors are grateful to Dr. Sandro Galea, Robert A. Knox Professor and Dean at the Boston University of Public Health, and to the National Center for Disaster Mental Health Research, for permission to use the data from the Galveston Bay Recovery Study. The work has been supported by a Discovery Grant to M. E. Thompson (RGPIN-2016-03688) from the Natural Sciences and Engineering Research Council of Canada.

References

- Anthopolos, R., Chen, Q., Sedransk, J., Thompson, M. E., Meng, G., & Galea, S. (2020). A Bayesian growth mixture model for complex survey data: clustering post-disaster PTSD trajectories (21 p.).
- Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1–59.
- Best, N., Richardson, S., & Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14, 35–39.
- Blanchard, E. B., Jones-Alexander, J., Buckley, T.C., et al. (1996). Psychometric properties of the PTSD Checklist (PCL). *Behavioral Research Therapy*, 34, 669–673.
- Eberly, L. E., & Carlin, B. P. (2000). Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in Medicine*, 19, 2279–2294.
- Gruebner, O., Lowe, S. R., Tracy, M., Joshi, S., Cerdá, M., Norris, F. H., Subramanian, S. V., & Galea, S. (2016a). Mapping concentrations of posttraumatic stress and depression trajectories following Hurricane Ike. *Scientific Reports*, 20166, 32242.
- Gruebner, O., Lowe, S. R., Tracy, M., Cerdá, M., Joshi, S., Norris, F. H., & Galea, S. (2016b). The geography of mental health and general wellness in Galveston Bay after Hurricane Ike: a spatial epidemiologic study with longitudinal data. *Disaster Medicine and Public Health Preparedness*, 10, 261–273.
- Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, 35, 1491–1523.
- Hartley, H. O., & Rao, J. N. K. (1962). Sampling with unequal probabilities and without replacement. *Annals of Mathematical Statistics*, 33, 350–374.
- Haziza, D., Mecatti, F., & Rao, J. N. K. (2008). Evaluation of some approximate variance estimators under the Rao-Sampford unequal probability sampling design. *Metron - International Journal of Statistics*, 66, 91–108.
- Leroux, B. G., Lei, X., & Breslow, N. (1999). Estimation of disease rates in small areas: a new mixed model for spatial dependence. In M. E. Halloran, & D. Berry (Eds.), *Statistical models in epidemiology, the environment and clinical trials* (pp. 135–178). New York: Springer Verlag.
- MacNab, Y. C. (2003). Hierarchical Bayes spatial modeling of small-area rates of non-rare disease. *Statistics in Medicine*, 22, 1761–1773.
- Pietrzak, R. H., Van Ness, P. H., Fried, T. R., Galea, S., & Norris, F. H. (2013). Trajectories of posttraumatic stress symptomatology in older persons affected by a large-magnitude disaster. *Journal of Psychiatric Research*, 47, 520–526.
- Rao, J. N. K., Verret, F., & Hidirolou, M. A. (2013). A weighted composite likelihood approach to inference for two-level models from survey data. *Survey Methodology*, 39, 263–282.
- Ribatet, M., Cooley, D., & Davison, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22, 813–845.
- Thompson, M. E., Sedransk, J., Fang, J., & Yi, G. Y. (2022). Bayesian inference for a variance component model using pairwise composite likelihood with survey data. *Survey Methodology*, 48, 73–93.
- University of Michigan Survey Research Center/Institute for Social Research. (2010). *The Galveston Bay Recovery Study: Report on Survey Procedure and Approach*.
- Valliant, R., Adams, T., & Wagner, J. (2009). Sample design documentation Galveston Bay recovery survey 2008–2009. *Survey Research Operations, Production Sampling Group, University of Michigan Survey Research Center*, 1–18.
- Yi, G. Y., Rao, J. N. K., & Li, H. (2016). A weighted composite likelihood approach for analysis of survey data under two-level models. *Statistica Sinica*, 26, 569–587.