# WEB MATERIAL

## WEB APPENDIX 1. IDENTIFICATION OF SUPERMARKET, PHYSICAL ACTIVITY VENUES AND FOOD OUTLET LOCATIONS

Supermarket business establishment data were acquired for the following years 2007, 2008, 2009, and 2010. The creation of the supermarket business establishment data that was used to generate the kernel density data followed a 6 step process: 1) records with addresses inside California with SIC codes corresponding to food retail stores such as grocers-retail, food markets, farm markets, fruits, vegetables, and produce-retail, and grocers-wholesale were selected from the database; 2) a word count of all words in the establishment name field was generated to guide the term search and recategorization process; 3) the records were also manually examined for common words and store names that corresponded with supermarkets and large grocery stores; 4) the above two methods generated a list of search terms for both inclusion and exclusion as a supermarket. Common supermarket chain names for inclusion included the following: Albertson, Lucky, Vons, Ralphs, Safeway, Whole Foods, General Hill, Trader Joe, Stater, Raley, Big Bear Market, Gelson, Bel Air, Food 4 Less, Albeco, Henrys Market, Wild Oats, Nob Hill, Nugget, Supervalue, Save-A-Lot, Save Foods, Fresh Easy, Grocery Outlet, Kroger, Save Mart, Food Max, Smart and Final, Pavilion, Vallarta, Cardenas Market, Foods Co, Win Co Food, 99 Ranch Market, Superior Grocers, Big Saver Foods, Best Way Supper, Payless Foods, Lion Food, Spencer Fresh Market, Valu Plus Food, Super Max Discount Food, Berkeley Bowl, Diablo Food, Mollie Stone, Pavilion, Ray Food Place, Sprouts Farm Market, Superior Warehouse, Tawa Supermarket, and Pak N Save Food. Other related terms such as Supermarket, Scolari Food, and Supermercado (Note this is only a selection) were also included but with restrictions placed on words before and after the search term; 5) The selected records were then further restricted to records that had an annual sales volume over $2 million and a square footage greater than 2,499; 6) The remaining records were then spot checked by store name and address in google maps and street view.

The physical activity data were created in a similar fashion to the supermarket data described above using a 3 step process: 1) Records with addresses inside California from a variety of SIC codes that corresponded to leisure and physical activity businesses were selected for inclusion; 2) these records were then cleaned to include and exclude establishment names based on their correspondence to leisure and physical activity. Word search terms were developed for each SIC category by manually examining establishment names. The following are a selection of terms used for inclusion: yoga, club, gym, fitness, YMCA, boys and girls club, rink, studio, cage, martial art, taekwondo, indoor, league, academy, school, youth, arena, pool, Pilates, golf, sailing, swim, lifting, athletic, tennis, volleyball, ball, fencing, bowl, horse, hockey, baseball, football, archery, boot camp, rock climb, water polo, recreation park, racing, personal training, pitching, ski club, softball, scouts, boxing, playground, program, dance, rugby, polo, racquet, and summer camp. The following are a selection of terms used for exclusion: shop, rent, sporting goods, supply, fishing, tour, authority, museum, launch, maintenance, motor, vehicle, outfitting,

lifeguard, aqua safe, and RV park. Please note both these word lists are not an exhaustive list of the terms used; 3) The remaining records were then spot checked by establishment name and address in google maps and street view where records not pertaining to leisure and physical activity were manually removed using their establishment name.

The unhealthful food outlet data were also created in a similar fashion to the supermarket data described above using a 3 step process: 1) Records with addresses inside California from SIC codes that corresponded to unhealthful food outlets were selected for inclusion in categories such as: grocers-retail, food markets, grocers-wholesale, service stations-gasoline and oil, food products-retail, discount stores, liquors-retail, beer and ale-retail, wines-retail, and convenience stores; 2) these records were then cleaned to include and exclude establishment names based on their correspondence to unhealthful food outlets. Word search terms were developed for each SIC category by manually examining establishment names. The following are a selection of terms used for inclusion to capture convenience and corner stores: 7 eleven, seven eleven, circle k store, am pm mini, quik stop, mini market, express mini mart, quick stop, convenience, mini, express, fast, ez stop, kwik, quick, beverage, stop n shop, on the run, snack, fast and easy, mart, lucky 7, lucky seven food, corner store, qwik korner, tower mart, and bonfare market. Please note this is not an exhaustive list of terms used; 3) The remaining records were then spot checked by establishment name and address in google maps and street view where records not pertaining to unhealthful food outlets were manually removed using their establishment name.

WEB TABLE 1. MISSING DATA IMPUTATION, KPNC DIABETES REGISTRY, NORTHERN CALIFORNIA, 2007-2011

With the exception of HbA1c and the Charlson comorbidity score, all other measures used in the analysis had non-missing values. Multiple linear regressions (using the *mi impute chained* command in STATA with 10 burn-in cycles) were used to impute HbA1c values (inverse-square transformed) and the Charlson comorbidity score. Individual age, gender and race/ethnicity (Non-hispanic white, Hispanic, Asian, Black), all model covariates and HbA1c and Charlson scores from other years were included as regressors in imputation models. Annual average BMI value and a kernel density measure for all businesses were included in imputation models as auxiliary variables. Imputed values were trimmed to be within the range of observed values and imputed Charlson scores were rounded to the nearest integer.

| | % Missing | | | |
|---|---|---|---|---|
| **Variable** | **2007** | **2008** | **2009** | **2010** |
| HbA1c | | | | |
| Pre-imputation | 14.9% | 16.2% | 21.6% | 24.4% |
| Post-imputation | 3.4% | 4.5% | 6.5% | 7.0% |
| Charlson comorbidity | | | | |
| Pre-imputation | 8.6% | 7.0% | 9.2% | 15.5% |
| Post-imputation | 2.5% | 3.0% | 2.8% | 4.8% |

WEB TABLE 2. MODEL RESULTS AND SENSITIVITY ANALYSES, KPNC DIABETES REGISTRY, NORTHERN CALIFORNIA, 2007-2011

| Model | # subjects (observations) | Supermarket gain $\beta_1$ | 95% CI[a] | Supermarket loss $\beta_2$ | 95% CI[a] |
|---|---|---|---|---|---|
| Model 0: Unadjusted | | | | | |
| A1c ≥ 9.0 | 18,343 (47,125) | 0.10 | 0.01, 0.21 | 0.18 | 0.08, 0.27 |
| A1c 8.0 - 9.0 | 18,013 (48,693) | 0.02 | -0.06, 0.09 | 0.11 | 0.03, 0.16 |
| A1c 6.5 - 8.0 | 75,123 (207,390) | 0.05 | 0.02, 0.08 | 0.01 | -0.02, 0.03 |
| A1c < 6.5 | 48,521 (133,513) | -0.02 | -0.03, 0.01 | 0.03 | 0.00, 0.05 |
| Model 1: Full model | | | | | |
| A1c ≥ 9.0 | 18,343 (47,125) | 0.10 | 0.01, 0.21 | 0.16 | 0.05, 0.23 |
| A1c 8.0 - 9.0 | 18,013 (48,693) | 0.01 | -0.08,0.09 | 0.12 | 0.03,0.16 |
| A1c 6.5 - 8.0 | 75,123 (207,390) | 0.05 | 0.02, 0.08 | 0.01 | -0.01, 0.03 |
| A1c < 6.5 | 48,521 (133,513) | -0.02 | -0.03, 0.01 | 0.02 | 0.00, 0.04 |
| Model 2: Model 1 + year effects | | | | | |
| A1c ≥ 9.0 | 18,343 (46,860) | 0.07 | -0.01, 0.19 | 0.08 | -0.01, 0.16 |
| A1c 8.0 - 9.0 | 18,013 (48,485) | 0.01 | -0.08,0.08 | 0.09 | 0.01, 0.14 |
| A1c 6.5 - 8.0 | 75,123 (206,535) | 0.04 | 0.02, 0.07 | -0.01 | -0.03, 0.01 |

| | | | | | |
|---|---|---|---|---|---|
| A1c < 6.5 | 48,521 (132,926) | -0.02 | -0.04,0.01 | 0.02 | -0.01, 0.04 |
| Model 3: Model1 + subject fixed effects | | | | | |
| A1c ≥ 9.0 | 18,343 (46,860) | 0.11 | 0.00, 0.25 | 0.20 | 0.06, 0.27 |
| A1c 8.0 - 9.0 | 18,013 (48,485) | 0.06 | -0.07, 0.13 | 0.11 | 0.02, 0.14 |
| A1c 6.5 - 8.0 | 75,123 (206,535) | 0.04 | 0.02, 0.09 | 0.01 | -0.01, 0.04 |
| A1c < 6.5 | 48,521 (132,926) | -0.02 | -0.04, 0.01 | 0.02 | -0.01, 0.05 |
| Model 4: Model 1 with complete case sample | | | | | |
| A1c ≥ 9.0 | 14,896 (30,374) | 0.07 | -0.03, 0.23 | 0.14 | 0.01,0.25 |
| A1c 8.0 - 9.0 | 15,470 (34,529) | 0.04 | -0.06, 0.13 | 0.06 | -0.03,0.15 |
| A1c 6.5 - 8.0 | 66,821 (156,964) | 0.07 | 0.03,0.10 | 0.01 | -0.02,0.03 |
| A1c < 6.5 | 42,341 (95,559) | -0.01 | -0.03, 0.02 | 0.03 | 0.00, 0.05 |
| Model 5: Model 1 by poverty rate strata | | | | | |
| Poverty ≥ 20% | 23,272 (61,612) | 0.03 | -0.02, 0.08 | 0.05 | -0.02, 0.09 |
| Poverty < 20% | 136,728 (373,194) | 0.02 | 0.01, 0.05 | 0.04 | 0.02, 0.06 |
| Model 6: Long difference (four-year) | | | | | |
| A1c ≥ 9.0 | 14,017 (14,017) | -0.06 | -0.13, 0.13 | -0.04 | -0.16, 0.08 |
| A1c 8.0 - 9.0 | 14,950 (14,950) | -0.04 | -0.11, 0.10 | 0.01 | -0.06, 0.11 |
| A1c 6.5 - 8.0 | 64,045 | 0.03 | -0.03, 0.05 | 0.02 | -0.02, 0.04 |

| | | | | | |
|---|---|---|---|---|---|
| | (64,045) | | | | |
| A1c < 6.5 | 40,958<br>(40,958) | 0.00 | -0.03, 0.03 | 0.00 | -0.03, 0.03 |

Model 0 : Unadjusted associations.

Model 1: Full model: adjusting for Charlson Comorbidity Index, diabetes medication use, PA density, unhealthful food outlet density, population density, poverty rate and median census block group housing value

Model 2: Full model with year fixed effects to account for potential secular changes in A1c across time.

Model 3: Full model with individual fixed effects such that subjects serve as their own controls not only with respect to their annual A1c levels but also with respect to their A1c trajectories over time.

Model 4: Full model excluding observations with any missing values to examine the impact of the data imputation process on estimates.

Model 5: Full model stratified by neighborhood poverty rate to explore whether the differences in estimates across A1c strata are attributable to underlying differences in neighborhood poverty.

Model 6: Long difference model to estimate the effect of supermarket change over a longer time horizon. All model variables are defined as change over four years (2007-2010). Patients who moved residence between 2007 and 2010 were excluded from the sample.

WEB TABLE 3. CHARACTERISTICS OF STUDY SAMPLE (IN 2007) BY BASELINE A1C STRATA KPNC DIABETES REGISTRY, NORTHERN CALIFORNIA

| | A1c < 6.5% (n =43,521) | A1c 6.5-8.0% (n = 68,358) | A1c 8.0-9.0% (n = 16,098) | A1c ≥ 9.0% (n = 15,175) |
|---|---|---|---|---|
| | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) |
| Individual-level variables | | | | |
| Age | 64.6 (12.7) | 62.6 (12.1) | 58.3 (11.6) | 54.6 (11.4) |
| % Female | 46.3% | 48.3% | 44.7% | 43.7% |
| Race | | | | |
| Asian | 14.0% | 20.6% | 19.3% | 16.4% |
| Black | 7.8% | 9.2% | 10.5% | 12.8% |
| Hispanic | 16.6% | 18.9% | 24.0% | 30.3% |
| White | 55.1% | 44.6% | 39.7% | 34.4% |
| Other | 6.4% | 6.7% | 6.5% | 6.1% |
| BMI | 30.9 (6.8) | 31.6 (7.1) | 32.5 (7.2) | 32.8 (7.4) |
| Comorbidity Score | 1.9 (1.3) | 1.8 (1.2) | 1.8 (1.2) | 1.7 (1.1) |
| % on insulin, TZDs, metformin or sulfonylureas | 44.2% | 72.8% | 82.5% | 80.8% |
| Neighborhood-level variables | | | | |
| % with supermarket presence | 53.2% | 54.2% | 53.9% | 55.3% |
| Fast food & convenience store kernel density | 1.1 (1.3) | 1.2 (1.2) | 1.2 (1.2) | 1.2 (1.2) |
| Physical activity venue density | 1.4 (1.6) | 1.5 (1.7) | 1.5 (1.7) | 1.5 (1.7) |
| Median home price | $521,511 ($223,486) | $513,376 ($215,911) | $493,040 ($210,805) | $481,162 ($203,680) |
| Poverty rate | 9.5% (9.5%) | 9.8% (9.6%) | 10.6% (10.1%) | 11.3% (10.5%) |
| Population density | 7,591 (7,863) | 8,155 (8,517) | 8,387 (8,792) | 8,727 (8,711) |

Differences across baseline A1c strata were significant at the p<0.001 level with the exception of race (p-value<0.05) and physical activity venue density (p-value <0.05).