

**A NEW METHOD FOR
ESTIMATING RACIAL/ETHNIC
DISPARITIES WHEN
SELF-REPORTED RACE/ETHNICITY
IS UNAVAILABLE**

*Marc N. Elliott
RAND
June 4, 2007
Academy Health*

Coauthors

This is joint, ongoing work with RAND colleagues

- Allen Fremont
- Nicole Lurie
- Peter A. Morrison
- Philip Pantoja
- Allan Abrahamse

Determining Race/Ethnicity

- To address health disparities, we need to classify populations by race/ethnicity (R/E)
- Self-report is considered the gold standard
- But self-reports are often unavailable and can take years to obtain
- Availability of R/E from other sources (e.g. CMS) limited

**Monitoring R/E Disparities
within Health Plans**

- We draw on research supporting efforts by health plans to monitor R/E disparities in health among their enrollees
- Their aims
 - Improve equity
 - Overall Quality Improvement
- Most plans collect enrollee's name & address, but not R/E

**Why Should Health Plans
Use Indirect Measures**

- Offers efficient and inexpensive way to:
 - Understand R/E composition & distribution of patients served
 - Estimate R/E disparities in care & outcomes
 - Fill gaps until sufficient self-reported R/E available
 - Assess potential bias in self-reported data
 - Examine contributing factors (e.g., SES)
- “Geocoding allows use of GIS mapping and decision tools”

**Current Name and Address
Methods**

- Surname lists with high sensitivity & specificity created for
 - Hispanics, by US Census Bureau
 - Asians, by Lauderdale & Kestenbaum
- Neighborhood Contextual Analysis
 - Uses local geographic information
 - Includes geo-coding of census information by residential address

Limitations of Surname Lists

- Ambiguous names limit sensitivity/specificity (e.g. Lee, Ohara)
- Spouses may adopt each other's surname upon marriage, or transfer to kids:
 - "John Smith" and "Mrs. Smith" (Maria Lopez)
 - "John & Maria Smith-Lopez"
 - "Soledad O'Brien" born to Irish & Hispanic parents
- Can't reliably distinguish Blacks from non-Hispanic whites

Limitations of Geo-coding

- If a census block group is
 - 90% black, 10% white, knowing someone's block group tells you a lot about their race/ethnicity
 - 25% black, 25% Hispanic, 25% White/Other, and 25% Asian, knowing someone's block group tells you little about their race/ethnicity
- The usefulness of geocoding varies regionally
- Typically, it is most informative for blacks

Logic of Our Hybrid Approach

- Integrate complementary strengths of surname lists & geo-coding
 - Surname lists best distinguishes Asians and Hispanics from others
 - Geocoding best distinguishes Blacks from others
- *How* you combine this information matters
 - Bayes' Theorem may be the best approach

Analogy to Medical Diagnostic Testing

Medical Diagnostic Testing	Indirect Estimation of R/E
Patient has prior probability of disease Base rate for the patient's risk group	Here we use the (4-category) race/ethnic distribution of the census block group
Apply diagnostic test with a given sensitivity and specificity for detecting the disease	Check whether name appears on each of 2 surname lists (with their Sen & Spe)
Bayes' Theorem turns the prior probability, sensitivity, and specificity into a posterior probability of disease.	-> posterior (combined) probabilities of belonging to each of 4 race/ethnic categories

Example 1: Block Group is 25% each Black, White, Asian, Hispanic

Probability of Given R/E	If On Asian List	If On Spanish List	If On Neither List
Asian	97%	2%	19%
Hispanic	1%	92%	8%
Black	1%	3%	36%
White/Other	1%	3%	36%

Example 2: Block Group is 67% White 11% Black, Asian, Hispanic

Probability of Given R/E	If On Asian List	If On Spanish List	If On Neither List
Asian	93%	1%	7%
Hispanic	1%	79%	3%
Black	1%	3%	13%
White/Other	6%	17%	78%

Comparing Selected Approaches to Obtaining R/E

- Our new *Bayesian Surname & Geocode approach (BSG)*
 - Uses posterior probabilities from surname and geocoding
 - is compared to:
- The *Classification with Surname and Geocoding (CSG)*
 - (1) Labels Hispanic if name on Spanish surname list; otherwise
 - (2) Labels Asian if the name on Asian surname list; otherwise
 - (3) If resides in a block group >= 66% Black, classified as black; otherwise
 - (4) Classified as non-Hispanic White
- *Census Prevalence only (CP)*
 - Uses geocoded r/e prevalence directly

Estimates of R/E Prevalence

N=1.9 million	Hispanic	Asian	Black	White/ Other	Overall Error vs. S-R
SR	8.9%	5.0%	8.0%	78.1%	(0)
BSG	10.0%	4.5%	9.1%	76.4%	1.6%
CP	10.8%	4.2%	9.0%	76.0%	2.0%
CSG	9.2%	2.9%	3.0%	84.9%	6.2%

Efficiency Compared to Geocoding Alone

n=1.9 million	Hispanic	Asian	Black	White/ Other	Weighted Average
BSG	260%	388%	115%	162%	174%
CSG	247%	365%	71%	131%	141%

Average Error in Disparity Estimates for 6 HEDIS Diabetes Measures

	Black vs. White	Hispanic vs. White
BSG	1.4%	1.8%
CP	2.7%	3.7%
CSG	8.2%	3.4%

Conclusions

- Surname analysis & geo-coding offer health plans powerful new ways to identify race/ethnicity from administrative records
- Our Bayesian approach efficiently integrates these sources
 - Doubles information from geocoding alone
- Because this information has geographic links, GIS opens the door to greater insight still