nielsen

IMPROVING THE ACCURACY OF BLOCK GROUP DATA FROM THE AMERICAN COMMUNITY SURVEY

Ken Hodges
May 29, 2014

BACKGROUND

- ACS data for block groups

    -- Concern over accuracy

    -- Large margins of error

    -- Conspicuous outliers

- ACS sample

    -- Smaller than census long form

    -- Many estimates based on few responses

BACKGROUND

Block Groups by Number of ACS Responses (2006-2010)

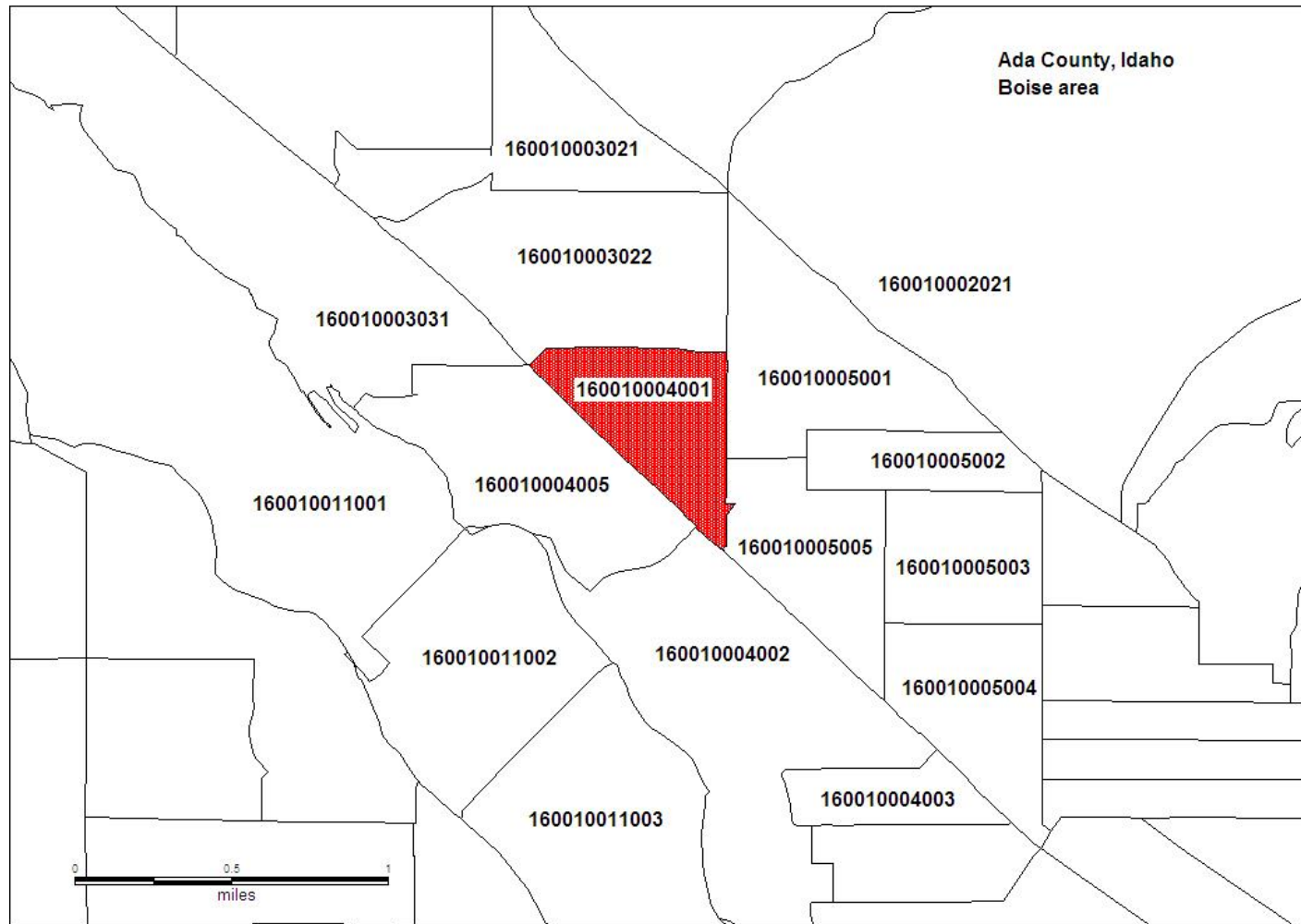| ACS Responses	N	Percent
Missing	380	0.2
Less than 10	2,650	1.2
10 to 19	18,264	8.4
20 to 49	132,669	60.9
50 to 99	53,315	24.5
100 to 199	9,555	4.4
200 to 499	896	0.4
500 or more	11	0.0
Total	217,740	100.0

BACKGROUND

- Accuracy better for aggregations

  -- But some dismiss value of block group data

  -- Critical of availability

- More productive to seek improvements

- Paper describes: **ACS Touch Method**

  -- Effort to improve accuracy

  -- Before use in Nielsen products

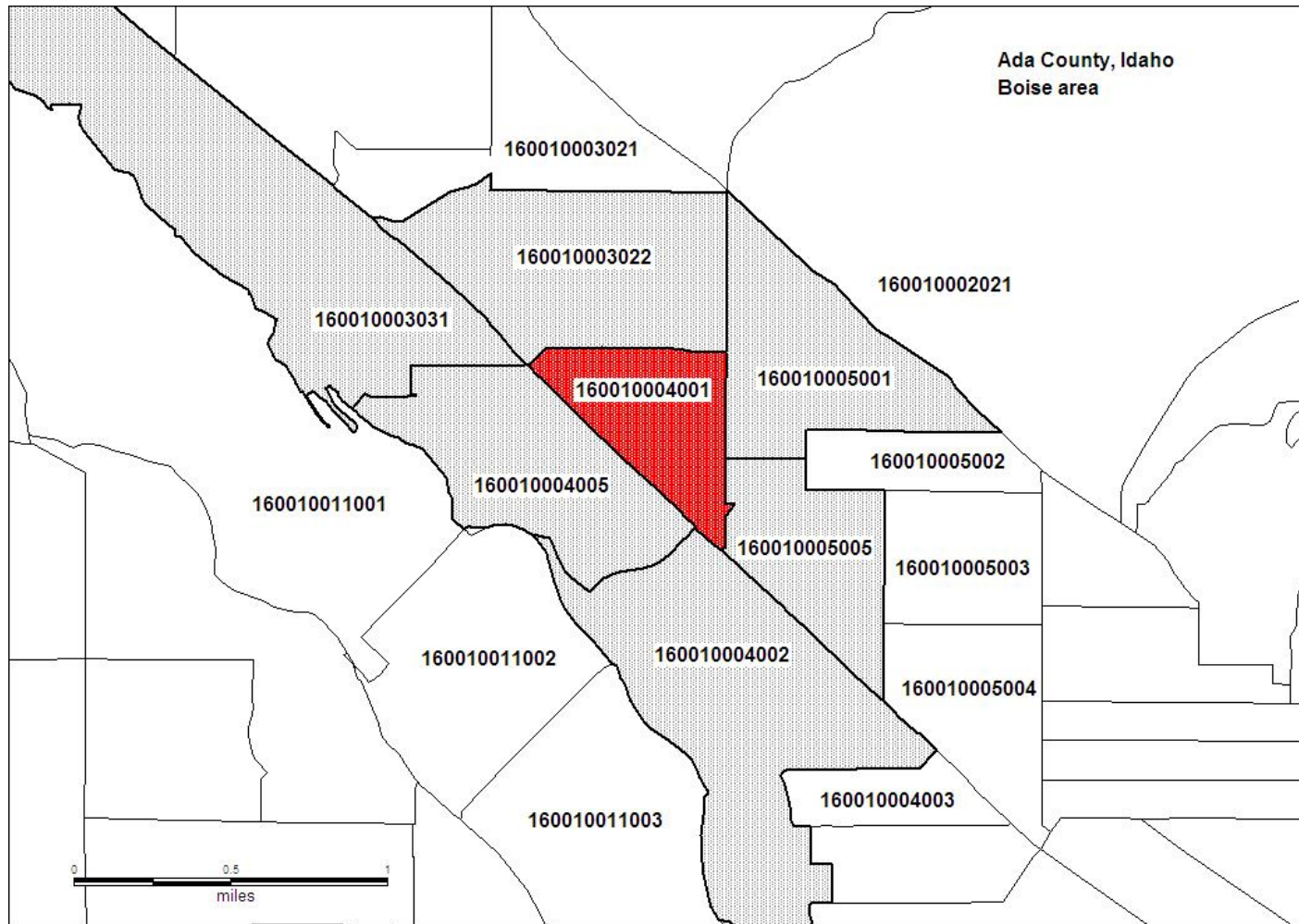 -- Also evaluation against 2010 census data

ACS TOUCH METHOD

- Effort to improve accuracy of ACS block group data

  -- Augment with data from adjacent block groups

  -- Expectation: adjacent BGs tend to be similar

  -- A form of "borrowing strength"

- Requires mega data processing

  -- But conceptually simple

  -- Easy to explain to clients

ACS TOUCH METHOD

Ada County, Idaho
Boise area

160010003021

160010003022

160010003031

160010002021

160010004001

160010005001

160010005002

160010011001

160010004005

160010005005

160010005003

160010011002

160010004002

160010005004

160010011003

160010004003

ACS TOUCH METHOD

ACS TOUCH METHOD

- For each ACS table and each block group

- Three distributions

  -- <u>ACS Published</u>: Data as published in ACS Summary File

  -- <u>ACS Touch</u>: Published combined with adjacent BGs

  -- <u>ACS Weighted</u>: Weighted avg of Published and Touch

- Weight based on Number of ACS responses

  -- If 100+ responses, Published weight = 100 pct

  -- Touch weight increased as responses decrease

2006-2010 ACS PERCENT DISTRIBUTION OF HOUSEHOLD TYPE BY HOUSEHOLD SIZE: BG 16 001 0004.00 1 IN IDAHO

HH Type and Size	ACS Published	ACS Touch	ACS Weighted
Total Households	328	4,082	328
Family 2 persons	36.9	20.1	23.3
Family 3 persons	8.5	9.8	9.6
Family 4 persons	21.3	12.5	14.2
Family 5 persons	3.7	3.1	3.2
Family 6 persons	0.0	1.9	1.6
Family 7+ persons	0.0	0.0	0.0
Nonfamily 1 person	3.4	39.7	32.8
Nonfamily 2 persons	4.0	8.9	7.9
Nonfamily 3 persons	0.0	2.2	1.7
Nonfamily 4 persons	0.0	0.0	0.0
Nonfamily 5 persons	0.0	0.0	0.0
Nonfamily 6 persons	0.0	0.0	0.0
Nonfamily 7+ persons	22.3	1.8	5.7
Unweighted HU	19		

EVALUATION

- Key question

  -- Does this approach make data more (or less) accurate?

- Evaluation

  -- Looked at "Households by Type and Size"

  -- Available for both ACS and census

  -- Compared all 3 ACS distributions vs. census

  - 2006-2010 data for all BGs

  - Index of dissimilarity

  -- An imperfect but revealing comparison

2006-2010 ACS PERCENT DISTRIBUTION OF HOUSEHOLD TYPE BY HOUSEHOLD SIZE: BG 16 001 0004.00 1 IN IDAHO

HH Type and Size	2010 Census	ACS Published	ACS Touch	ACS Weighted
Total Households	376	328	4,082	328
Family 2 persons	22.3	36.9	20.1	23.3
Family 3 persons	17.6	8.5	9.8	9.6
Family 4 persons	7.7	21.3	12.5	14.2
Family 5 persons	5.1	3.7	3.1	3.2
Family 6 persons	1.9	0.0	1.9	1.6
Family 7+ persons	0.5	0.0	0.0	0.0
Nonfamily 1 person	32.2	3.4	39.7	32.8
Nonfamily 2 persons	11.4	4.0	8.9	7.9
Nonfamily 3 persons	1.1	0.0	2.2	1.7
Nonfamily 4 persons	0.3	0.0	0.0	0.0
Nonfamily 5 persons	0.0	0.0	0.0	0.0
Nonfamily 6 persons	0.0	0.0	0.0	0.0
Nonfamily 7+ persons	0.0	22.3	1.8	5.7
Index of Dissimilarity		**50.4**	**15.2**	**14.4**

EVALUATION

- OK . . . the big outlier is improved

- What about across all block groups?

Mean Index of Dissimilarity for All Block Groups:

Alternative ACS Distributions versus 2010 Census

ACS Comparison	Mean IOD
ACS Published vs. Census	19.4
ACS Touch vs. Census	12.2
ACS Weighted vs. Census	12.2

- Does this differ by N of ACS responses?

MEAN INDEX OF DISSIMILARITY FOR BLOCK GROUP ACS DISTRIBUTIONS OF HOUSEHOLD TYPE BY SIZE BY NUMBER OF ACS RESPONSES

ACS Responses	N	Published	Touch	Weighted
All Block Groups*	216,598	19.4	12.2	12.2
Fewer than 10	1,900	60.3	32.4	32.1
10 to 19	18,255	29.3	15.0	14.8
20 to 29	50,291	23.2	13.1	13.0
30 to 39	48,379	19.4	12.1	12.0
40 to 49	33,996	16.9	11.5	11.4
50 to 59	22,114	15.3	11.1	11.0
60 to 69	13,884	14.2	10.7	10.9
70 to 79	8,512	13.4	10.4	11.1
80 to 89	5,314	12.7	10.2	11.3
90 to 99	3,491	12.2	10.0	11.7
100 or more	10,462	10.8	9.2	10.8

* Block groups with 2010 households greater than 0.

EVALUATION

- "Touch" better than "Published" even where 100+ ACS responses

  -- Better to use ACS Touch for all BGs?

- Another view

  -- How often do Touch and Weighted improve over Published?

PERCENT OF BLOCK GROUPS WHERE ACS TOUCH AND WEIGHTED HAD LOWER INDEX OF DISSIMILARITY THAN ACS PUBLISHED: HOUSEHOLD TYPE AND SIZE BY NUMBER OF ACS RESPONSES

ACS Responses	N	Touch	Weighted
All Block Groups*	216,598	82.5	88.4
Fewer than 10	1,900	91.2	92.1
10 to 19	18,255	92.9	95.3
20 to 29	50,291	89.4	94.0
30 to 39	48,379	85.2	92.8
40 to 49	33,996	80.5	91.8
50 to 59	22,114	77.3	91.3
60 to 69	13,884	74.2	91.6
70 to 79	8,512	72.0	92.4
80 to 89	5,314	70.4	93.5
90 to 99	3,491	68.2	93.0
100 or more	10,462	65.4	0.0

* Block groups with 2010 households greater than 0.

EVALUATION

- OK . . . improvement across BGs

- But is there a negative impact at aggregate levels?

Mean Index of Dissimilarity for ACS Distributions: Block Groups Summed to County and National Levels

ACS Comparison	County	National
ACS Published vs. Census	4.8	2.0
ACS Touch vs. Census	4.6	1.8
ACS Weighted vs. Census	4.7	2.1

- Conclusion: No major negative impact on aggregate areas

CONCLUSIONS

- Comprehensive test not feasible

  -- ACS is only source for most data it provides

  -- ACS period estimates not comparable with census

- But limited evaluation suggests potential

  -- ACS outliers improved

  -- Mean IODs are reduced

  -- Improved accuracy in most BGs

  -- Accuracy of aggregations preserved

CONCLUSIONS

- Could debate merits of Touch vs. Weighted

- But the potential for improvement is clear

- And the application could be refined

  -- More selective designation of "touch" BGs

  -- Those similar on decennial characteristics

- The choice is ours . . .

  -- Dismiss BG data as error-prone

  -- Or seek to reduce error, and maximize value

Thank You !

Ken.hodges@nielsen.com