Strategies for Increasing Efficiency of Cellular Telephone Samples

Kurt Peters[1], William Robb[1], Cristine Delnevo[2], Daniel A. Gundersen[2]

Prepared for:

FedCASIC

March 2014

[1] **ICF International**
[2] **Rutgers School of Public Health**

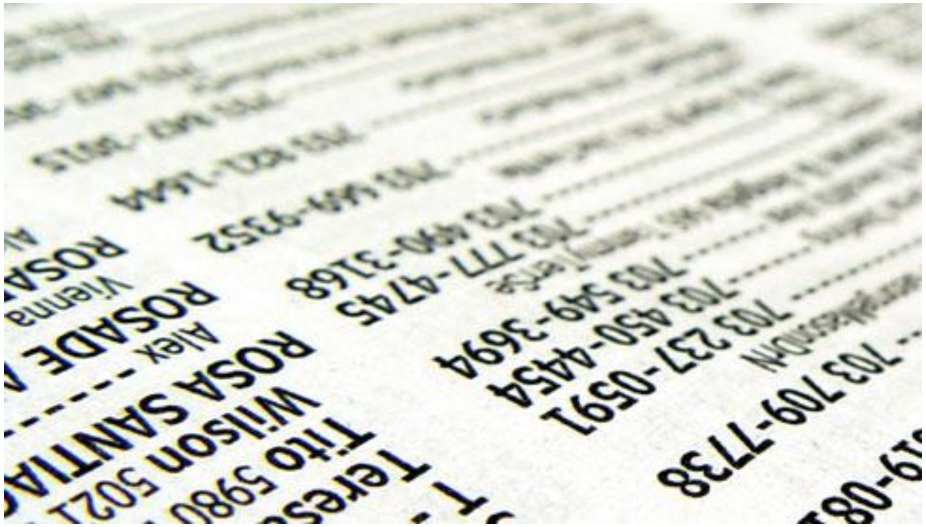
Contact: kurt.peters@icfi.com

Overview

- **A study of cell phone (CP) sample flags assessed the potential for increased efficiency**
  - The study is based on a national random digit dial (RDD) sample of CP numbers used to conduct interviews with young adults

Overview

- **Two sample flags appended by vendor (MSG) were examined:**
  - A Cell-WINS indicator designed to identify active CP numbers
  - A billing ZIP code



Overview

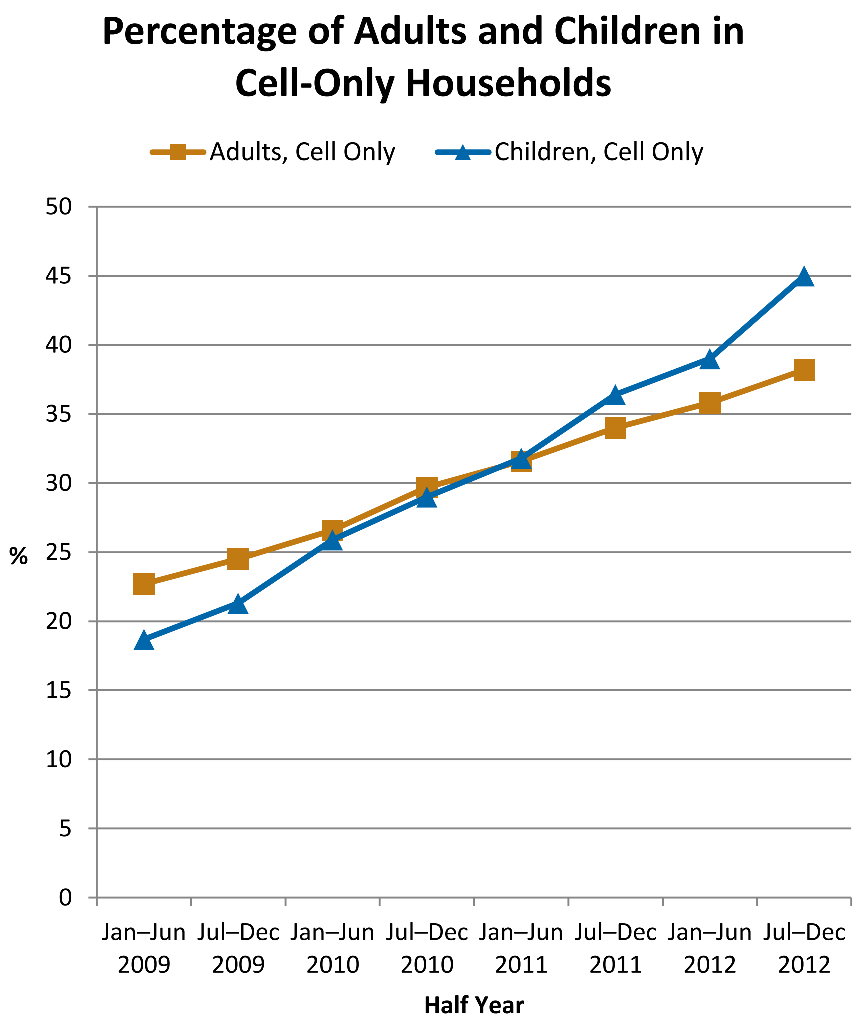
- **Tests revealed Cell-WINS to be an accurate indicator of active phone status for CPs**
  - This may make it tempting to use only “active” sample for RDD CP surveys
  - However, our research suggests doing so may introduce coverage bias
- **Billing ZIP code less accurate**
  - But may be useful for targeting broader geographies

National Young Adult Health Study (NYAHS)

- **National representation**
- **RDD cell phone frame**
- **Screen for adults ages 18 – 34**
- **Collects data on smoking trends in young adult population in support of prevention efforts**
- **Fielded from 1 August 2013 – 1 January 2014**

Cell Phone Usage

- **45% of children and 36.5% of adults lived in cell-only households as of Dec 2012**
  - Health status and health insurance measures differ between landline and cell phone households
- **Increasingly important to cover cell-only population**
  - How to do this efficiently in an RDD design?

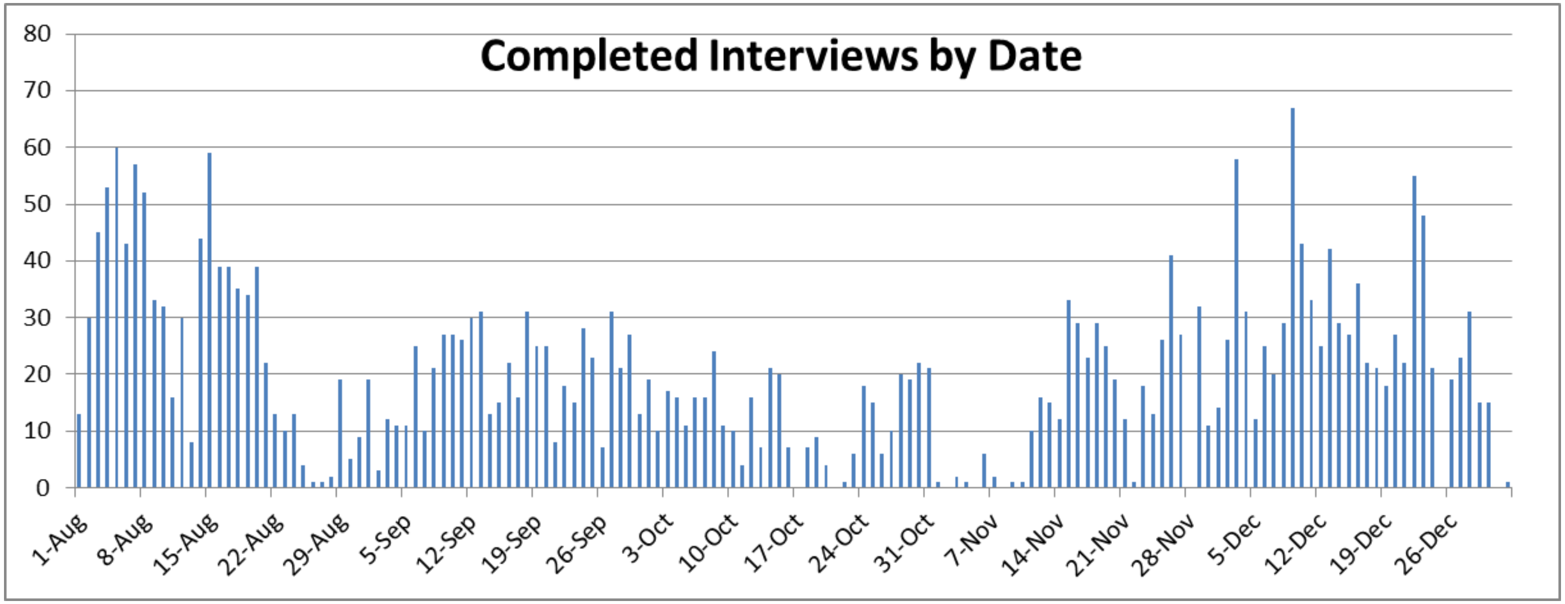


Source: National Health Interview Survey

Methodology & Initial Results

NYAHS Sample

- National Random Digit Dial (RDD) Cell Phone Sample
- 205,732 numbers drawn
- 3,095 completed interviews



Sample Flags

- **Cell-WINS flag for active CPs**
  - MSG: “A real-time, non-intrusive screening process that accurately identifies inactive telephone numbers within a Cellular RDD sample”
- **Billing ZIP Code**
  - Appends the ZIP code associated with the billing address for the phone number

Source: http://www.m-s-g.com/Web/genesys/cell-wins.aspx

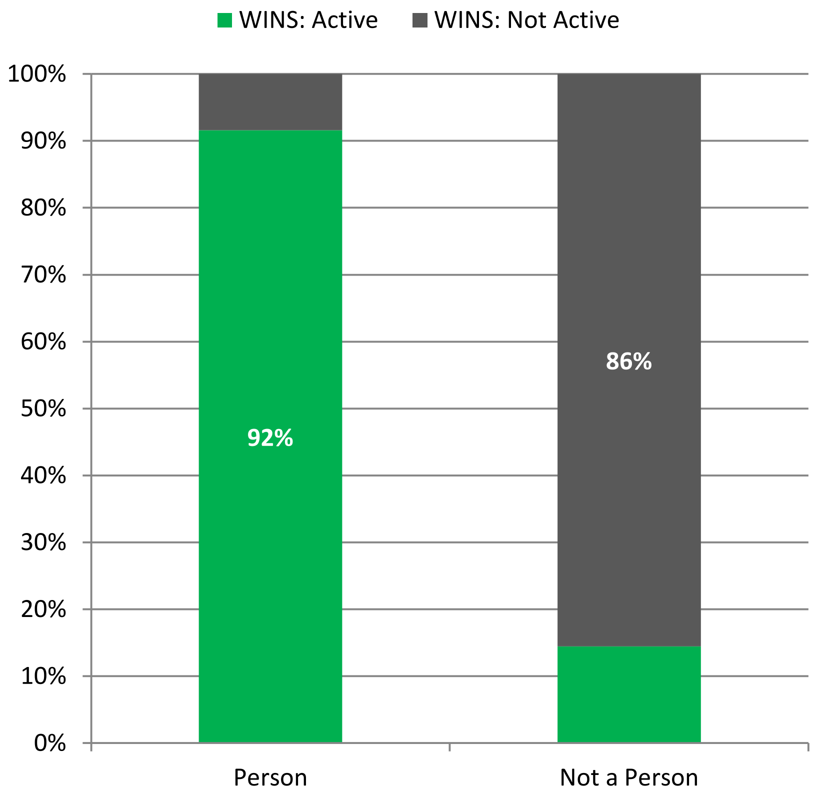
The Experiment

- **Sample put through both the Cell-WINS and ZIP-append flagging process**
- **205,413 CP numbers dialed using a 6-attempt protocol**
  - These records were used to assess the accuracy of Cell-WINS and the appended billing ZIP code
- **To assess productivity, sample was separated by study for a portion of the calling**
  - Productivity = Completes / Hour
  - Standard shift reporting collected data on the number of completes and the number of interviewer hours per shift over 141 shifts (26 August-23 September)

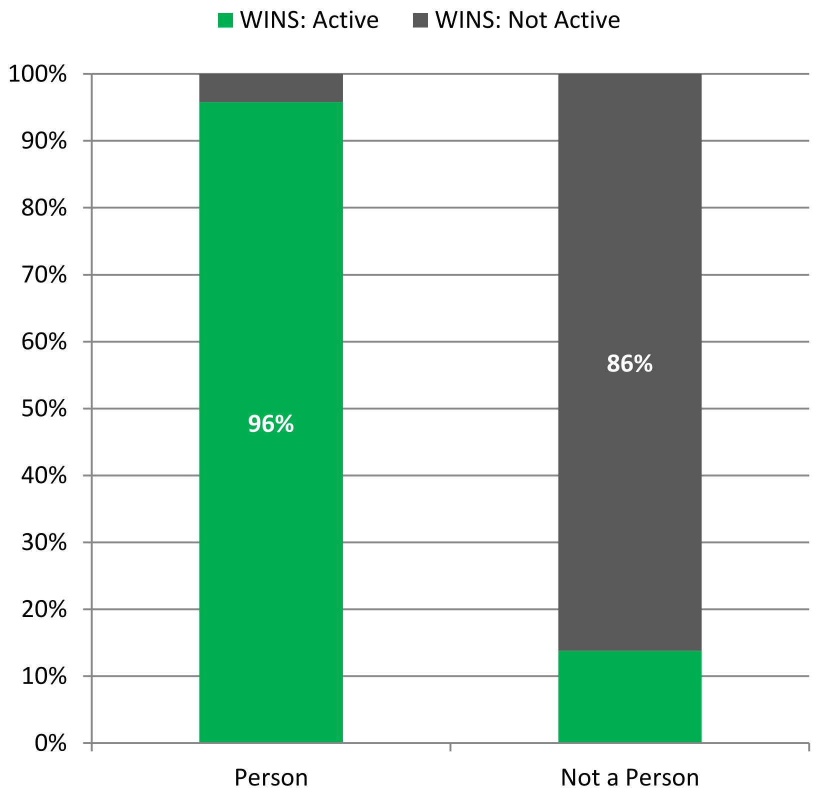
Accuracy

Cell-WINS Accuracy

All Records

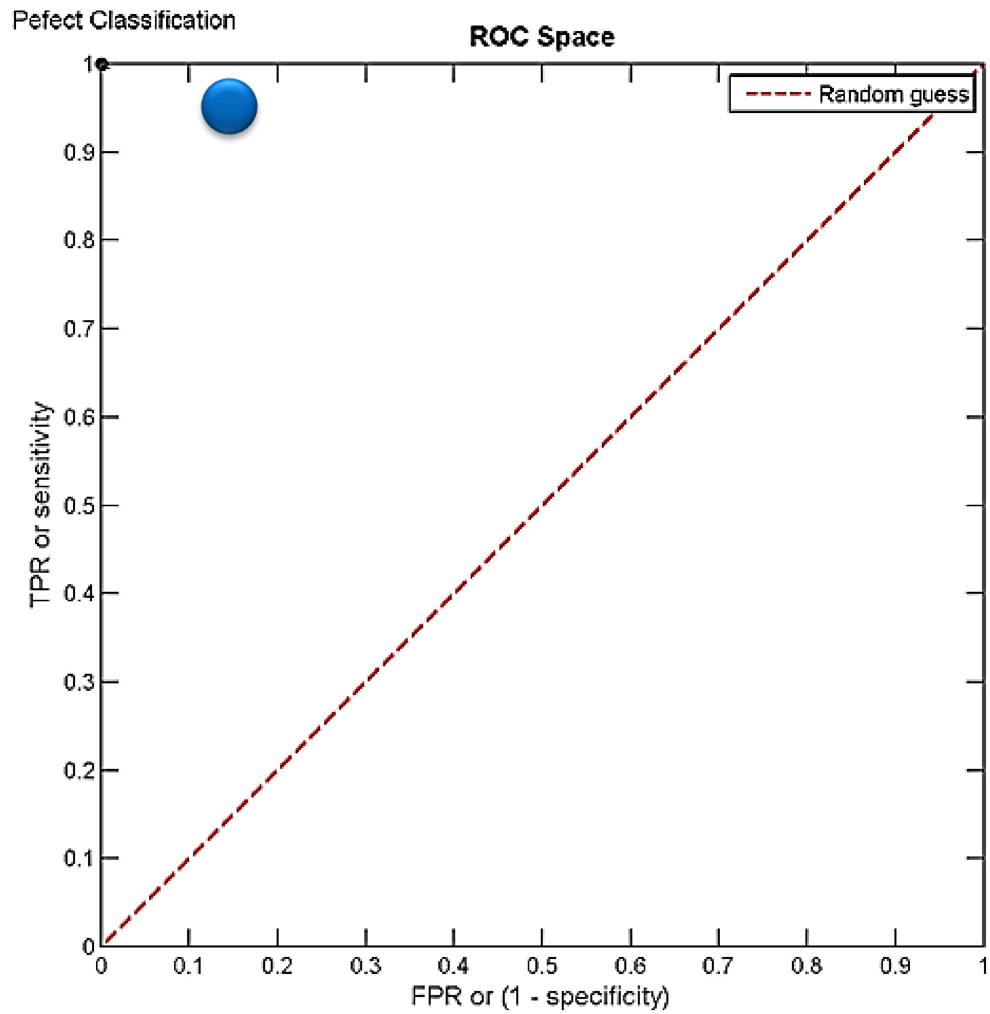


Excluding Unresolved



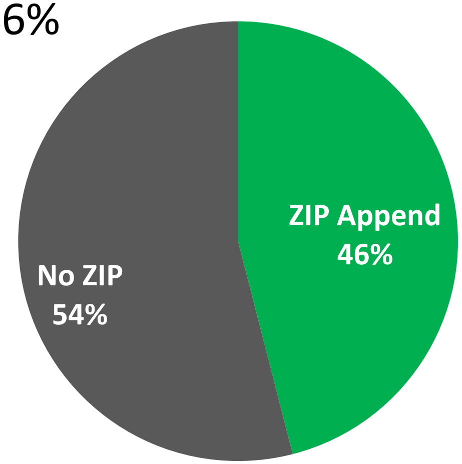
Cell-WINS Accuracy

- **Excluding unresolved records:**
  - **True Positive Rate** = 96%
  - **True Negative Rate** = 86%
  - **False Positive Rate** = 14%
  - **False Negative Rate** = 4%



Billing ZIP Code

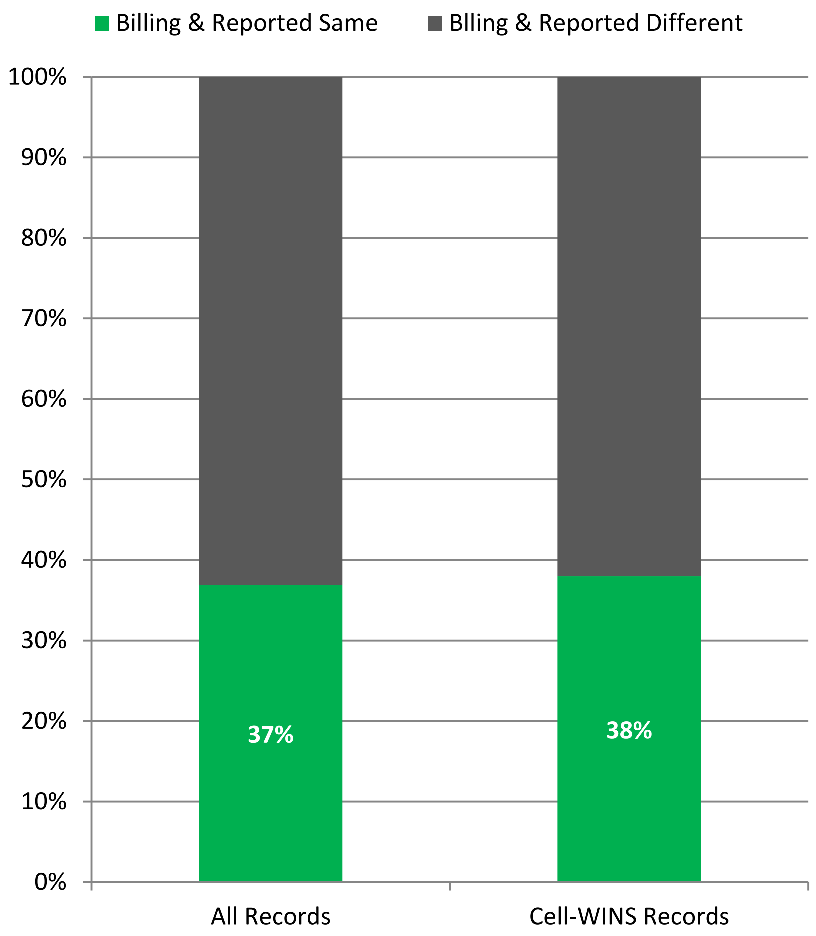
- **First assigned when phone is purchased**
- **Follows person as they move (assuming they get the bill at residence)**
  - Note that Rate Centers do not update when phone moves
  - For example, one author's billing ZIP code is Union City, NJ, but his rate center is South Burlington, VT, where he bought the first phone associated with that number
- **Not all sampled records match to a billing zip code**
  - Overall append rate for this study = 46%



Billing ZIP Code Accuracy

- **For records with an appended ZIP that resulted in a complete, we computed the match rate against self-reported ZIP (*N* = 1,287)**
  - No interaction with Cell-WINS
  - Dutwin (2014) found similar results in an analysis of appended billing ZIP (31% match rate)

Overall match rate =
46% * 37% =
17%

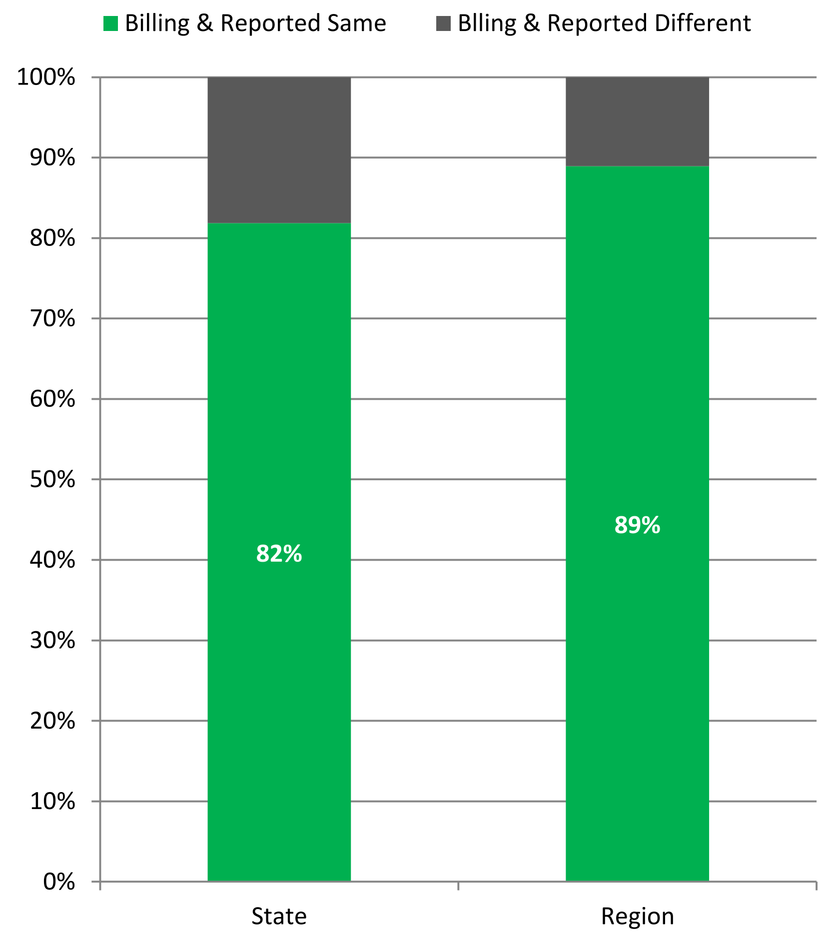


Dutwin, D. (2014). ***Cellular telephone methodology: Present and future*****. AAPOR Webinar.**

Billing ZIP Code Accuracy

- **Accuracy improves as geography broadens out**
  - Billing ZIP may be useful for geographic targeting, especially at broader geographies
  - But low append rate still requires a “no billing ZIP” stratum to restore lost coverage

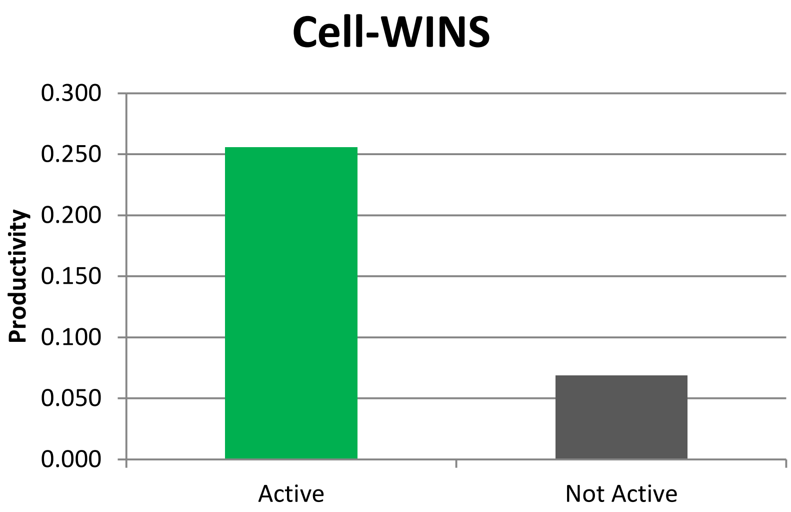
Overall match rate
State = 38%
Region = 41%

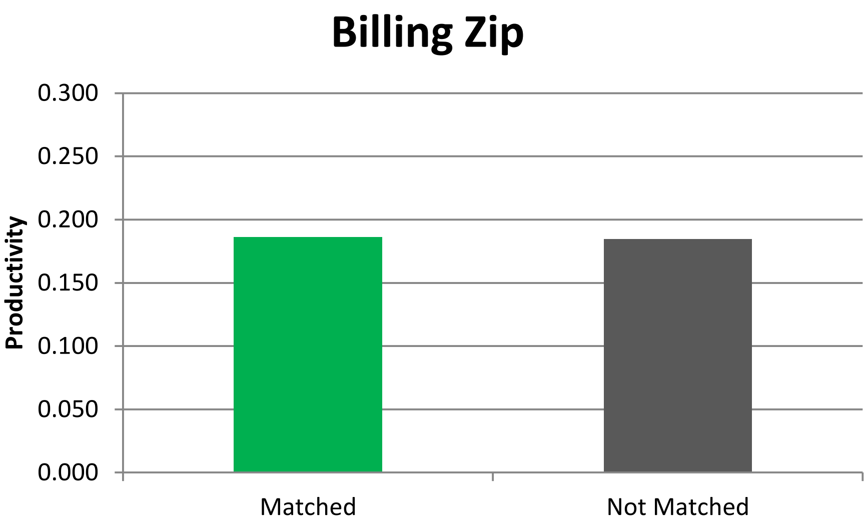


Productivity

Productivity

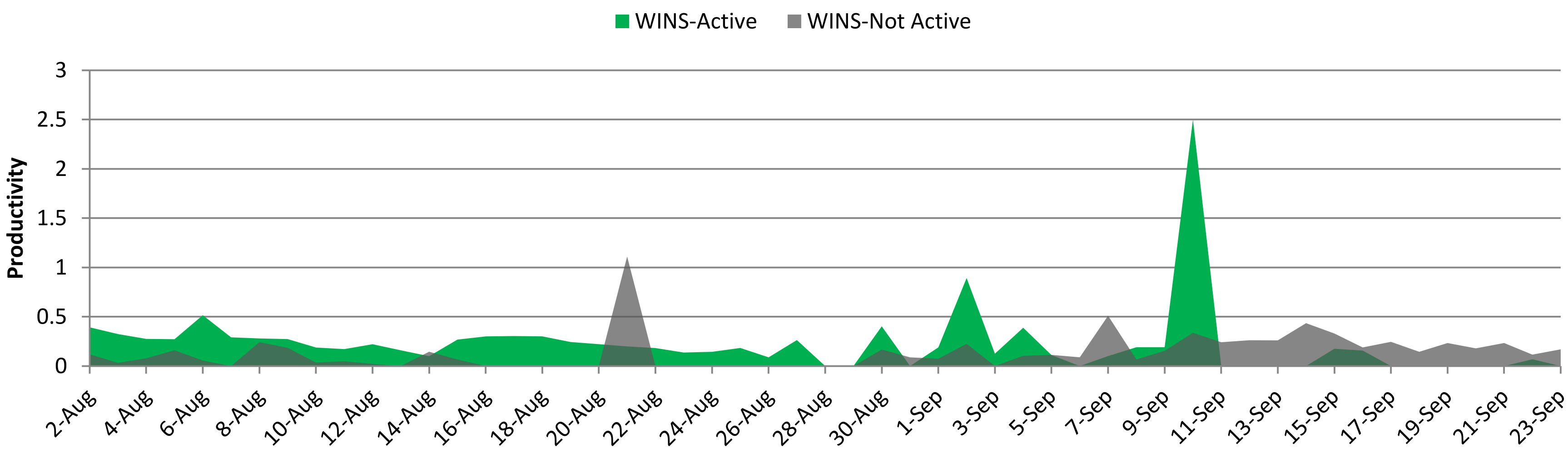
- **Productivity defined as completes per hour**
  - Computed from shift-level call center data
  - Productivity was higher for Cell-WINS sample, but not for Billing ZIP sample





Productivity

- **Modeled productivity as a function of Cell-WINS and Billing ZIP**
  - $Pr \sim WINS + ZIP + WINS \times ZIP$
- **Model R^2 = .04, p = .086**
  - Productivity data exhibit high variability, so the large observed average differences were masked
- **Even if not statistically significant, the average difference for Cell-WINS is of operational significance**



Bias Analysis

Bias Analysis: Cell-WINS

- **Key NYAHS items were compared between Cell-WINS Active vs. Non-Active**

Item	Odds Ratio (Non-Active vs. Active)
Current smoker	1.8
100+ cigarettes in lifetime	1.8
Use smokeless tobacco	1.9
CP is a smartphone	0.5
Have healthcare coverage	0.5
Unemployed/Looking	2.1
Minority	1.8
HH Income <= $25K	2.8
Educational attainment	0.4

CONCLUSION

Cell-WINS Non-Active sample is demographically different: less healthy, less employed/educated, higher minority, lower SES

Note: All differences significant, $p < .05$

Bias Analysis: Cell-WINS

- **Key NYAHS items were compared between Billing ZIP missing vs. appended**

Item	Odds Ratio (Missing vs. Appended)
CP is a smartphone	0.7
Have healthcare coverage	0.8
Enrolled in college prev 6 mos	1.3
Unemployed/Looking	1.5
Minority	2.0
Hispanic	1.5
HH Income <= $25K	1.6
Educational attainment	0.5

CONCLUSION
Billing ZIP-Missing sample is demographically different: similar to Cell-WINS sample (lower SES) but not as strongly skewed

Note: All differences significant, $p < .05$

Using Cell-WINS for Cell Phone Oversampling

Cell-WINS Oversampling

- **Cell-WINS Active sample was about 3.7 times more productive than Not Active sample**
  - However, clear demographic differences exist between these two groups
  - Dialing only Cell-WINS Active sample would introduce substantial coverage bias

Cell-WINS Oversampling

- **Our solution was to oversample Cell-WINS Active records**
  - Analogous to density stratification of list-assisted landline RDD sample
- **Optimal allocation proportions were determined following Cochran's (1977) formula:**

$$n_h = \frac{N_h \tilde{S}_h / \sqrt{C_h}}{\sum (N_h \tilde{S}_h / \sqrt{C_h})}$$

- **Where**
  - $N_{Active} = 62$ (based on 62% of sample flagged as active)
  - $N_{Inactive} = 38$ (based on 38% of sample flagged as not active/unknown)
  - $\tilde{S}_{Active} = 0.85$, averaged across *SD* for 6 sentinel variables
  - $\tilde{S}_{Inactive} = 0.96$, averaged as above
  - $C_{Active} = \frac{1}{Pr_{Active}} = 4.15$
  - $C_{Inactive} = \frac{1}{Pr_{Inactive}} = 26.32$

Cell-WINS Oversampling

- **The resulting optimal allocation is 78.4% to Cell-WINS Active (vs. Not Active)**
  - Oversampling factor $= \frac{78.4}{21.6} = 3.6$
  - Expected DEFF due to weighting $= (\sum_h W_h w_h)(\sum_h W_h / w_h) = 1.6$

Conclusions

- **Cell-WINS flag**
  - Very accurate (96% TPR, 86% TNR)
  - Population **mis**categorized as not active is demographically different (lower SES)
  - Oversampling strategy is recommended to balance efficiency with coverage
- **Billing ZIP append**
  - Baseline append rate is low (46%)
  - Accuracy against self-reported ZIP is low (37%), but higher for state/region (82%/89%)
  - May be useful for oversampling at broader geographies, but low append rate and demographic differences require coverage of a “No Billing ZIP” stratum

Thank You!

Contact: kurt.peters@icfi.com

ICF INTERNATIONAL