Appendix MM

English MDOT GPS Household Letter
(for GPS packets)

«FIRSTNAME» «LASTNAME»
«ADDRESS»
«CITY», «STATE» «ZIP»-«ZIP4» «DATENOW»

Dear «FIRSTNAME»,

Thank you for choosing to participate in the MI Travel Counts survey!

The information you provide will help ensure that future transportation projects reflect what your community needs. Remember that we value your input, no matter how much or how little you travel.

NEXT STEPS:

1. Record your travel information using the GPS devices and Travel Logs.

- Use the **Travel Logs** to record all places visited by your household on «FIRSTTRAVELDAY».
- Use your **GPS devices** from **«FIRSTTRAVELDAY» to «LASTTRAVELDAY»**. GPS equipment is being provided for each household member between the ages of 16 and 75 as noted in the table below. Please follow the instructions for use that are provided in the package.

Person	Name	Age	GPS Unit ID
«PERSON1»	«NAME1»	«AGE1»	«GPSUNITID1»
«PERSON2»	«NAME2»	«AGE2»	«GPSUNITID2»
«PERSON3»	«NAME3»	«AGE3»	«GPSUNITID3»
«PERSON4»	«NAME4»	«AGE4»	«GPSUNITID4»
«PERSON5»	«NAME5»	«AGE5»	«GPSUNITID5»
«PERSON6»	«NAME6»	«AGE6»	«GPSUNITID6»

- Return the **GPS devices** and the completed **Equipment Usage Log** in the pre-paid FedEx packaging immediately after your travel period. Refer to the return instructions in the package. ***Be sure to keep your Travel Logs for the next step.***

2. Once your GPS equipment has been returned, we will contact you via text, email, or phone and ask you to confirm your travel information in one of the following ways:

- **Online:** Go to **www.MITravelCounts.com.**

  Click "Report Travel" and **enter your PIN#: «PINNO»**.

- **By Phone:** Call 1-855-774-1800 to talk with a Westat study team member.

Once we confirm the travel information for all household members, we will send your **$«INCENTIVE»** check for participating in the survey. Remember, participation is voluntary and your personal information will be kept confidential, as required by law.

Thank you again for helping to move Michigan forward!

Appendix NN

English GPS Equipment Usage Log

Equipment Usage Log

Complete this form to help you keep track of your participation in the study. After Day 3, return it to us along with your GPS equipment.

Instructions: Fill in the appropriate columns (**with a Yes or No**) at the end of each travel day.

		Day 1 «DOW» «ASSN»			Day 2 «TD+1» «NEXTDATE»		Day 3 «TD+2» «DATEAFTERNEXT»	
	PERSON:	**Traveled?**	**Use GPS?**	**Filled in Travel Log?**	**Traveled?**	**Use GPS?**	**Traveled?**	**Use GPS?**
1	«NAME1»							
2	«NAME2»							
3	«NAME3»							
4	«NAME4»							
5	«NAME5»							
6	«NAME6»							
7	«NAME7»							
8	«NAME8»							

«HHID»-«GFLAG»

Remember…

To thank you for your participation, you will receive $25 for each person in your household that carries a GPS device, as long as every person in the household:

- ✓ Completes their Travel Log on the assigned travel date: <<**TRAVELDATE**>>
- ✓ Uses their GPS device as instructed
- ✓ Returns their GPS devices after the last GPS day
- ✓ When contacted, reports their travel either online or over the phone

Your check for $«INCENTIVE» will be mailed once travel and GPS information is processed and confirmed for everyone in your household.

Appendix OO

English GPS Device Usage and Return Instructions

GPS Device Use Instructions

This package contains a GPS logging device for each person in your household between the ages of 16 and 75. Please refer to this document for directions about how to use and return the GPS equipment.

TURNING ON THE GPS DEVICE

1. Press and hold down the **silver** power button in the upper right-hand corner for about 5 seconds.
2. When the device turns on, all three lights will flash (**green**, **blue** and **red**). The **green** and **red** light will stay on.
3. Throughout the day, check to make sure the device is on. The **red light** should always be on, and the **green light** should be on (either solid or flashing). If these lights are not on, press and hold the **silver** power button to turn the device off and then turn it back on.

WHEN TO WEAR THE GPS DEVICE

You should wear the GPS device beginning on your assigned travel date and continuing through all days of your GPS study period (as listed on the letter in this package). You should turn on the GPS device every morning and wear the GPS device whenever you travel outside of your home.

HOW TO WEAR THE GPS DEVICE

When walking, biking, or riding public transportation, you should wear the GPS device on your waist or clipped to your bag or purse. If you are riding inside a vehicle such as a car or truck, you can continue to wear the GPS device on your waist or place your bag or purse on the seat. Make sure to keep the device with you at all times.

CHARGING THE GPS DEVICE

It is very important to charge the GPS device every night after using it. To charge the device:

1. Plug the larger end of the enclosed cable into the side of the GPS device.
2. Connect the opposite end of the cable into the wall plug adaptor and plug the adapter into the wall.
3. If the connection is right, the bottom light on the GPS device will light up in **red** indicating that it is charging. The **red** light may go off once the device is fully charged.

Be sure to recharge the GPS device every night!

Questions?

www.MITravelCounts.com

1-855-774-1800

Equipment Return Instructions
(See Other Side)

GPS Equipment Return Instructions

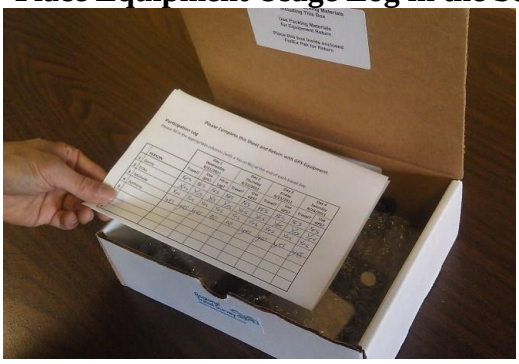
As soon as possible after your last travel day, collect all GPS devices and equipment, place them in the packaging material and box in which they arrived, and place the box inside the pre-paid FedEx Pak (and seal the Pak). Please include the completed Equipment Usage Log, but **remember to keep the Travel Logs** to confirm your travel once we contact you.

PACKAGING THE DEVICE FOR RETURN

Step 1:
Repackage the GPS devices

Step 2:
Place GPS devices into the box

Step 3:
Place Equipment Usage Log in the box



Step 4:
Place the box into the FedEx package

FEDEX RETURN OPTIONS

Option 1: Take the package to a FedEx Drop Box or to FedEx Office Location. To find the nearest location, visit www.fedex.com, or call 1-800-GO-FEDEX (1-800-463-3339).

Option 2: Schedule a FedEx pickup at your home or office. To schedule a pickup, visit www.FedEx.com/pickup, or call 1-800-GO-FEDEX (1-800-463-3339). Tell the FedEx representative you have a pre-paid return envelope.

Option 3: Have us schedule a FedEx pickup for you. Just call us at 1-855-774-1800.

REMEMBER, IN ORDER TO RECEIVE YOUR PARTICIPATION CHECK:

1. Use the enclosed GPS devices and Travel Logs
2. Fill out the Equipment Usage Log
3. Return GPS devices and the completed Equipment Usage Log in the FedEx package
4. Use your Travel Logs to confirm your travel information by web or phone

Equipment Use Instructions
(See Other Side)

Appendix PP

English GPS Equipment Retrieval Letter

«FIRSTNAME» «LASTNAME»
«ADDRESS» «SUITE»
«CITY», «STATE» «ZIP»

7/17/2015

«DATENOW»

Dear «FIRSTNAME» «LASTNAME»,

Thank you for your participation in the GPS portion of the MI Travel Counts survey!

Our records indicate that we have not received your GPS device(s) and we need your help to get the unit(s) back. Please return the equipment in the box in which it was sent, using the pre-paid FedEx envelope provided in the box. If you need a replacement envelope, please call **1-855-774-1800** to arrange for another envelope to be sent. In order to process your participation payment, we need to receive the device(s) from you.

If you have already returned the equipment, please contact our office at **1-855-774-1800** to ensure we received your device(s).

Thank you again for helping to move Michigan forward!

Laura Wilson
MI Travel Counts GPS Equipment Manager

Appendix QQ

Spanish GPS Household Letter
(for GPS Packets)

Estimado/a John,

¡Gracias por aceptar participar en la encuesta MI Travel Counts!

La información que proporcione ayudará a garantizar que los futuros proyectos de transporte reflejen lo que su comunidad necesita. Recuerde que valoramos su información, sin importar lo mucho o poco que viaje.

SIGUIENTES PASOS:

1. Registre su información de viajes usando los dispositivos GPS y los registros de viajes.

- Use los r**egistros de viajes** para registrar todos los lugares visitados por su familia en «FIRSTTRAVELDAY».
- Use sus **dispositivos GPS** desde **«FIRSTTRAVELDAY» hasta «LASTTRAVELDAY»**. Se está entregando equipo GPS para cada miembro de su familia entre las edades de 16 y 75 años, como se indica en la siguiente tabla. Siga las instrucciones de uso incluidas en el paquete.

Persona	Nombre	Edad	Identificación de la unidad GPS
«PERSON1»	«NAME1»	«AGE1»	«GPSUNITID1»
«PERSON2»	«NAME2»	«AGE2»	«GPSUNITID2»
«PERSON3»	«NAME3»	«AGE3»	«GPSUNITID3»
«PERSON4»	«NAME4»	«AGE4»	«GPSUNITID4»
«PERSON5»	«NAME5»	«AGE5»	«GPSUNITID5»
«PERSON6»	«NAME6»	«AGE6»	«GPSUNITID6»

- Devuelva los **dispositivos GPS** y el r**egistro de utilización del equipo** completado en el sobre con franqueo pagado de FedEx inmediatamente después de su período de viaje. Consulte las instrucciones de devolución en el paquete. ***Asegúrese de conservar los registros de viajes para el siguiente paso.***

2. Luego que ha sido devuelto su equipo GPS, nos comunicaremos con usted por mensaje de texto, correo electrónico o teléfono y le pediremos confirmar su información de viajes en una de las siguientes formas:

- **En línea:** Vaya a **www.MITravelCounts.com**.

  Haga clic en "Report Travel" (Informar los viajes) e **ingrese su n.º de PIN: <<PIN>>**.

- **Por teléfono:** Llame al teléfono 1-855-774-1800 para conversar con un miembro del equipo de estudio de Westat.

Luego que confirmemos la información de viajes de todos los miembros de la casa, le enviaremos su cheque por participar en la encuesta. Recuerde, la participación es voluntaria y su información personal se mantendrá en forma confidencial, como lo exige la ley.

¡Gracias nuevamente por ayudar al progreso de Michigan!

Appendix RR

Spanish GPS Device Use and Return Instructions

Instrucciones de uso del dispositivo GPS

Este paquete contiene un dispositivo de registro GPS para cada persona en su casa entre las edades de 16 y 75 años. También se han proporcionado registros de viajes a todos los miembros de la casa.

ENCENDER EL DISPOSITIVO GPS

- Para encender el dispositivo GPS, pulse y mantenga presionado el botón de energía **plateado** ubicado en la esquina superior derecha por aproximadamente cinco segundos. Cuando el dispositivo se enciende, todas las tres luces (**verde**, **azul** y **roja**) destellarán y las luces **verde** y **roja** deben permanecer encendidas.
- Asegúrese que el dispositivo esté encendido cada mañana y cuando esté fuera de casa. La **luz roja** siempre debe estar encendida, y la **luz verde** debe estar encendida (en forma continua o destellando). Si estas luces no están encendidas, pulse y mantenga presionado el botón de energía **plateado** para apagar el dispositivo, y luego pulse este botón y manténgalo presionado nuevamente para volver a encender el dispositivo.

CUÁNDO Y CÓMO USAR EL DISPOSITIVO GPS

- Debe usar el dispositivo GPS cada vez que viaje fuera de su hogar empezando en su fecha de viaje asignada y continuando durante todos los días de su período del estudio de GPS (como se indica en la carta incluida en este paquete).
- Al caminar, montar bicicleta o tomar el transporte público debe usar el dispositivo GPS en su cintura o enganchado en su bolso o cartera. Si está viajando en un vehículo, tal como un automóvil o camión, puede seguir usando el dispositivo GPS en su cintura o puede colocar su bolso o cartera en el asiento.
- La **luz verde** destellará cuando la información se esté recopilando.

CARGAR EL DISPOSITIVO GPS

- Cada noche después de usar cada dispositivo GPS, conecte un extremo del cable incluido (el extremo con el conector más largo) en el cable en el lado del dispositivo GPS. Conecte el extremo opuesto del cable en el adaptador del enchufe de la pared y conecte el adaptador en la pared. Si la conexión es correcta, la luz inferior en el dispositivo GPS se encenderá de color **rojo** indicando que el dispositivo se está cargando. La luz **roja** puede apagarse luego que el dispositivo esté completamente cargado.

Recargue el dispositivo GPS cada noche.

¿Tiene preguntas?
Llame al teléfono
1-855-774-1800
de lunes a viernes
de 9 a.m. a 9 p.m.
(hora del este)

Instrucciones para devolver el equipo
(Consulte el otro lado)

Instrucciones para devolver el equipo GPS

Tan pronto como sea posible después del último día de viaje, junte todos los dispositivos GPS, cables y adaptadores CA, colóquelos en el material de empaque y caja en que llegaron, y coloque la caja dentro del sobre con franqueo pagado de FedEx (y séllelo). Incluya el **Registro de utilización del equipo** completado, pero recuerde **conservar los Registros de viajes para confirmar su viaje cuando nos comuniquemos con usted**.

EMPACAR EL DISPOSITIVO PARA DEVOLUCIÓN

Paso 1:
Vuelva a empacar los dispositivos GPS

Paso 2:
Coloque los dispositivos GPS en la caja

Paso 3:
Coloque el Registro de utilización del equipo en la caja

Paso 4:
Coloque la caja en el sobre de FedEx

OPCIONES DE DEVOLUCIÓN A FEDEX

Opción 1: Lleve el paquete a un buzón de FedEx o a la ubicación de una oficina de FedEx. Para encontrar la ubicación más cercana, visite www.fedex.com, o llame al teléfono 1-800-GO-FEDEX (1-800-463-3339).

Opción 2: Programe el recogido de FedEx en su hogar u oficina. Para programar un recogido, visite www.FedEx.com/pickup, o llame al teléfono 1-800-GO-FEDEX (1-800-463-3339). Dígale al representante de FedEx que tiene un sobre de devolución con franqueo pagado.

Opción 3: Permítanos programar para usted un recogido de FedEx. Solamente llámenos al teléfono 1-855-774-1800.

RECUERDE, PARA RECIBIR SU CHEQUE DE PARTICIPACIÓN, DEBE:

1. Usar los dispositivos GPS y registros de viajes incluidos
2. Llenar el registro de utilización del equipo
3. Devolver los dispositivos GPS y el registro de utilización del equipo completado utilizando el embalaje de devolución de FedEx
4. Usar sus registros de viajes para confirmar su información de viajes por Internet o teléfono.

Instrucciones de uso del equipo
(Consulte el otro lado)

Appendix SS

Spanish GPS Equipment Retrieval Letter

«FIRSTNAME» «LASTNAME»
«ADDRESS»
«CITY», «STATE» «ZIP»-«ZIP4»

«DATENOW»

Estimado/a «FIRSTNAME» «LASTNAME»,

¡Gracias por su participación en la sección GPS de la encuesta MI Travel Counts!

Nuestros registros indican que no hemos recibido sus dispositivos GPS y necesitamos su ayuda para recuperar las unidades. Le agradeceríamos que devuelva el equipo en la caja en la que se envió, utilizando el sobre con franqueo pagado de FedEx que se proporcionó en la caja. Si necesita un sobre de reemplazo, llame al teléfono **1-855-774-1800** para ordenar el envío de otro sobre. Para procesar el pago por su participación necesitamos que nos envíe los dispositivos.

Si ya devolvió el equipo, comuníquese con nuestra oficina llamando al teléfono **1-855-774-1800** para asegurar que recibimos sus dispositivos.

¡Gracias nuevamente por ayudar al progreso de Michigan!

Laura Wilson
Gerente de equipos GPS de MI Travel Counts

Appendix TT

Arabic GPS household Letter
(for GPS Packets)

عزيزي (الاسم الأول)،

شكرا لموافقتك على المساهمة في استطلاع إحصاءات ميشيغان للرحلات!

ستساعدنا المعلومات التي تقدمها في ضمان ان مشروعات النقل المستقبلية ستعكس احتياجات مجتمعك بشكل صحيح. تذكر اننا نقدر لك ما تقدمه من مساهمات، بغض النظر عن كثرة رحلاتك أو قلتها.

الخطوات التالية:

1. **سجل معلومات رحلاتك باستخدام أجهزة تحديد المواقع وكذلك باستخدام سجلات الرحلات.**

- استخدم **سجلات الرحلات** لتسجيل الأماكن التي زارها أفراد اسرتك المعيشية خلال "يوم السفر الأول" «FIRSTTRAVELDAY».
- استخدم **أجهزة تحديد المواقع** بدءا من «FIRSTTRAVELDAY» إلى «LASTTRAVELDAY». يتم توفير أجهزة تحديد المواقع لكل عضو في الأسرة يتراوح عمره بين 16 إلى 75 عام كما هو موضح بالجدول أدناه. من فضلك اتبع إرشادات الاستخدام الموجودة في الرزمة.

فرد 1	الاسم:	السن	رقم تعريف الجهاز
«PERSON1»	«NAME1»	«AGE1»	«GPSUNITID1»
«PERSON1»	«NAME1»	«AGE1»	«GPSUNITID2»
«PERSON1»	«NAME1»	«AGE1»	«GPSUNITID3»
«PERSON1»	«NAME1»	«AGE1»	«GPSUNITID4»
«PERSON1»	«NAME1»	«AGE1»	«GPSUNITID5»
«PERSON1»	«NAME1»	«AGE1»	«GPSUNITID6»

- قم بإعادة **أجهزة تحديد المواقع** و**سجل استخدام المعدات** المكتمل مستخدما في ذلك رزمةFedEx للإعادة بمجرد انتهاء فترة رحلاتك. راجع إرشادات الإعادة الموجودة في الرزمة. **تأكد من احتفاظك بسجل الرحلات لاستخدامه في الخطوة التالية.**

2. **سوف نتصل بك، بمجرد تمام إعادة المعدات، عبر الرسائل النصية أو البريد الإلكتروني او الهاتف لنطلب منك تأكيد معلومات رحلاتك بواحدة من الطرق التالية:**

- **عبر شبكة الإنترنت: اذهب إلى الموقع:** www.MITravelCounts.com. **اضغط على "أبلغ عن الرحلات"** "Report Travel" **وأدخل الرقم السري الخاص بك:** «PINNO».
- **عبر الهاتف:** اتصل بالرقم 1-855-774-1800 لكي تتحدث مع أحد أعضاء فريق بحوث ويستات.

وبمجرد الانتهاء من تأكيد معلومات السفر لكل أعضاء أسرتك، سنرسل شيكا بقيمة دولار نظير مساهمتكم في الاستطلاع. تذكر أن المشاركة تطوعية وسيتم الحفاظ على سرية معلوماتك الشخصية، كما ينص على ذلك القانون.

شكرا لك مرة اخرى على مساعدتك في الدفع بميشيغان إلى الأمام!

Appendix UU

Arabic GPS Device Use and Return Instructions

«الاسم الاخير» «الاسم الاول»
«العنوان»»

«التاريخ الان» «الرمز البريدي 4»-«الرمز البريدي» «الولاية» ،«المدينة»

عزيزي (الاسم الأول) (الاسم الأخير)،

شكرا لك على مشاركتك في قسم جهاز تحديد المواقع من استطلاع إحصاءات ميشيغان **للرحلات!**

تشير سجلاتنا إلى أننا لم نتلقى أجهزة تحديد المواقع الخاصة بك ونود أن تساعدنا بأن تعيد لنا ذلك الجهاز (الأجهزة). رجاءً قم بإعادة المعدات مستخدما الصندوق التي وصلتك فيه ومظروف FedEx مسبق الدفع الموجود في الصندوق. في حالة احتياجك إلى مظروف بديل نرجو منك أن تتصل بالرقم **1-866-436-7828** لترتيب إرسال مظروف آخر لك. نحن نحتاج إلى ان نتلقى الأجهزة منك لكي نتمكن من إتمام خطوات إرسال مقابل الاشتراك المالي الخاص بك.

إذا كنت قد أعدت المعدات فعلا فاتصل من فضلك بمكتبنا على الرقم **1-866-436-7828** لمساعدتنا في تحديد موقع أجهزتك.

شكرا لك مرة اخرى على مساعدتك في الدفع بميشيغان إلى الأمام!

لورا ويلسون
مديرة معدات أجهزة تحديد المواقع في إحصاءات ميشيغان للرحلات

Appendix VV

Arabic GPS Equipment Retrieval Letter

«الاسم الاخير» «الاسم الاول»
«العنوان»»

«التاريخ الان» «الرمز البريدي 4»-«الرمز البريدي» «الولاية» ،«المدينة»

عزيزي (الاسم الأول) (الاسم الأخير)،

شكرا لك على مشاركتك في قسم جهاز تحديد المواقع من استطلاع إحصاءات ميشيغان **للرحلات!**

تشير سجلاتنا إلى أننا لم نتلقى أجهزة تحديد المواقع الخاصة بك ونود أن تساعدنا بأن تعيد لنا ذلك الجهاز (الأجهزة). رجاءً قم بإعادة المعدات مستخدما الصندوق التي وصلتك فيه ومظروف FedEx مسبق الدفع الموجود في الصندوق. في حالة احتياجك إلى مظروف بديل نرجو منك أن تتصل بالرقم **1-866-436-7828** لترتيب إرسال مظروف آخر لك. نحن نحتاج إلى ان نتلقى الأجهزة منك لكي نتمكن من إتمام خطوات إرسال مقابل الاشتراك المالي الخاص بك.

إذا كنت قد أعدت المعدات فعلا فاتصل من فضلك بمكتبنا على الرقم **1-866-436-7828** لمساعدتنا في تحديد موقع أجهزتك.

شكرا لك مرة اخرى على مساعدتك في الدفع بميشيغان إلى الأمام!

لورا ويلسون
مديرة معدات أجهزة تحديد المواقع في إحصاءات ميشيغان للرحلات

Appendix WW

Trip Mode and Purpose Imputation Procedure Memo

Westat®

An Employee-Owned Research Corporation®

Date: February 26, 2016

From: Marcelo Simas, Westat

Subject: MDOT HTS Trip Mode and Purpose Imputation Procedure

Introduction

This document provides information on how the GPS-derived place data collected as part of the MDOT Statewide Household Travel Survey (HTS) were augmented with attributes such as: travel mode and trip purpose. This process was applied to the 2,156 GPS households.

The final processed and imputed GPS-derived places cover three days of travel. Within this structure each travel day starts with a place whose arrival time is set to 3am and ends with a place with a departure time of 2:59am. The variable **travday** contains a number which indicates the travel day number to which a place belongs, with one (1) corresponding to the prompted recall (PR) day.

The following sections detail how various place attributes were imputed by combining place information derived from the GPS data with person and household data collected as part of the recruit interview and other ancillary data sources.

Travel Mode

Initial estimates for travel mode were generated using the procedure described in Simas Oliveira, et al. (2011)[1], these were then reviewed by analysts using information collected during the recruit interview on individual's usage and availability of modes such as school bus, bike and transit. During this initial review, analysts set trips which were done using a private vehicle to the driver

[1] Simas Oliveira, M. G. S. et al., 2011. *GPS-Assisted Prompted Recall Household Travel Survey to Support Development of Advanced Travel Model in Jerusalem, Israel.* Transportation Research Record 11-1166, pp. 16-23.

mode (5). Transit trips were reviewed using information on local bus and rail lines, which was displayed on the map along with GPS data, to select the appropriate travel modes between the multiple transit choices available in the survey. A similar quality check process was run on all trips associated with school bus and bike modes where an analyst reviewed them in TripBuilder, where information on the reported use of these modes captured in recruit was available.

Following this analyst review, the resulting travel modes were run through an automated rule-based procedure which relied on shared travel information to set private vehicles modes to either driver or passenger. A driver was selected between the members in the group with priority given to older persons, only those people who reported being a driver in the recruit data were eligible to be considered drivers. Once a person was set to a driver in a shared travel event he or she was also set to be the driver every time that they took place in group travel using a private vehicle mode. The final result was saved to the **mode** variable.

Trip Purpose

Decision trees were developed to impute trip purpose, in a process similar to the one described in Simas Oliveira (2014)[2]. Decision trees can be interpreted as a graph or flow chart in which internal nodes encapsulate a Boolean test, with each branch representing an outcome of this test and each leaf node representing a final classification label. A path from the root of the tree to a leaf node describes the classification rules. Trip purpose decision trees were created using the C50[3] package in the R[4] statistical tool. C50 is an open source implementation of the C4.5 algorithm used by Simas Oliveira (2014)[5].

[2] Simas Oliveira, M. G. S. et al., 2014. Evaluating Two Methods for Identifying Trip Purpose in GPS-based Household Travel Surveys. Washington, Transportation Research Record 14-3407, pp 33-41.

[3] Max Kuhn, Steve Weston, Nathan Coulter and Mark Culp. C code for C5.0 by R. Quinlan (2015). C50: C5.0 Decision Trees and Rule-Based Models. R package version 0.1.0-24. URL https://CRAN.R-project.org/package=C50

[4] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[5] Simas Oliveira, M. G. S. et al., 2014. Evaluating Two Methods for Identifying Trip Purpose in GPS-based Household Travel Surveys. Washington, Transportation Research Record 14-3407, pp 33-41.

Given the known limitations and predicting non-mandatory purposes identified in Simas Oliveira (2014)[6] a decision was made to first aggregate the reported purposes into the categories identified in Table 1.

Table 1: Aggregate Trip Purposes

Aggregate Purpose	Original Purpose Codes
Home activities	
Home	1, 2
Work, school & volunteer	
Work	3, 4, 5
School/studying	6
Volunteering	7
Shopping & errands	
Shopping	8, 9
Maintenance	10, 11, 12, 13
Social activities	
Discretionary	14, 15, 16, 17, 18
Travel-related activities	
Drop off/pick up passenger	19
Change travel mode/transfer	20

A decision tree was built for each of the six person types identified in Table 2. Only non-home locations were included in the tree models. All places which matched the home location on days two and three were set to purpose 1 (typical home activities). First and last places of days two and three were also set to purpose 1 (typical home activities).

Input variables for the decision tree building algorithm included person and place level variables, along with a range of computed and spatial variables. The trees were built using a confidence factor of 0.25 and at least 10 observations per leaf. The confidence factor determines how closely the tree conforms to the training set, and 0.25 is the default in C4.5. The 10-observations-per-leaf setting keeps the trees from being overly specific (i.e. avoids over-fitting).

Table 2: Person Lifecycle Categories

Id	Category	Description
1	Worker	Person is a full-time worker.
3	University Studend	Person is 18 years of age or older and a student.
4	Non-Worker	Person is 18 years of age or older and does not work nor goes to university (persons who were missing key attributes and could not be classified into the other categories were treated as non-workers).
5	Retiree	Person is retired.
6	Driving Age School Child (16-17yrs)	Person is between 16 and 17 years of age and goes to high-school.

[6] Simas Oliveira, M. G. S. et al., 2014. Evaluating Two Methods for Identifying Trip Purpose in GPS-based Household Travel Surveys. Washington, Transportation Research Record 14-3407, pp 33-41.

The reported cases for "Something else" (97) were made unavailable during the tree estimation process, the final training data set contained 8,785 records.

A series of computed variables were added to the place and person data records. This included classifying household members into one of eight life cycle categories listed in Table 2. Instead of using land use variables clusters for the aggregate non-mandatory and non-escorting purposes (shopping, maintenance, volunteer, discretionary and change of mode) were created by processing places across all GPS-PR households. The clustering process was also implemented in R using the DBSCAN[7] algorithm, with a tolerance of 50 meters. Dummy 1/0 variables were created to indicate when a place was near (distance < 150 meters) to one of these clusters. For escorting and work purposes two variables were computed (escorting_points and work_points) based on the proximity of a place to any place from the same person with either a escorting or work purpose on the PR day. A complete list of the computed variables and their definitions is included in Appendix A.

The variables used on each of the six decision trees were determined iteratively by examining the frequency distributions of the estimated purposes against the frequency distribution of the reported purposes. Appendix B contains detailed information on each of these trees along with estimated error rates by purpose and tree.

A confusion matrix, also referred to as a prediction-success table (PST) in travel forecasting, was constructed using the reported and estimated aggregated purposes for the PR day. This matrix shows actual choices as rows and modeled outcomes as columns, correct classifications appear on the matrix's diagonal. Within the context of a confusion matrix Type I errors are the sum of each column without the diagonal value, and likewise the Type II errors are sum of each row without the diagonal value.

Table 3: Confusion Matrix using Aggregated Purposes for the PR Day.

Reported/ Modeled	Change of Mode	Discretionary	Escorting	Home	Maintenance	School	Shopping	Volunteer	Work	Type II Error
Change of Mode	111	11	15	16	10	2	7	1	14	41%
Discretionary	1	1859	69	220	224	13	231	17	51	31%
Escorting	4	81	920	55	73	3	69	8	97	30%
Home	0	0	0	8303	0	0	0	0	0	0%

[7] Michael Hahsler (2015). dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms. R package version 0.9-6. https://CRAN.R-project.org/package=dbscan

Maintenance	10	125	40	66	2699	1	269	2	41	17%
School	1	27	28	59	10	163	2	3	31	50%
Shopping	5	83	17	0	390	1	1738	0	13	23%
Volunteer	3	13	20	9	10	2	3	105	16	42%
Work	6	131	59	123	141	5	89	14	2433	19%
Type I Error	21%	20%	21%	6%	24%	14%	28%	30%	10%	

Looking at Table 3, the purposes which were incorrectly selected more often (Type I Error) were Volunteer and Maintenance. The two reported purposes that most often showed different model results (Type II Error) were School and Volunteer. The non-home purposes which were correctly identified most often were maintenance, work and shopping, which showed match rates of 83%, 81% and 77%.

The six aggregate trip purpose decision trees, as well as the home purpose assignment rules, were applied to the 44,800 GPS places for days two and three. The results were saved to the delivery variable tpurp_agg. Figure 1 shows the final frequency distributions of the aggregate trip purposes by travel day number.

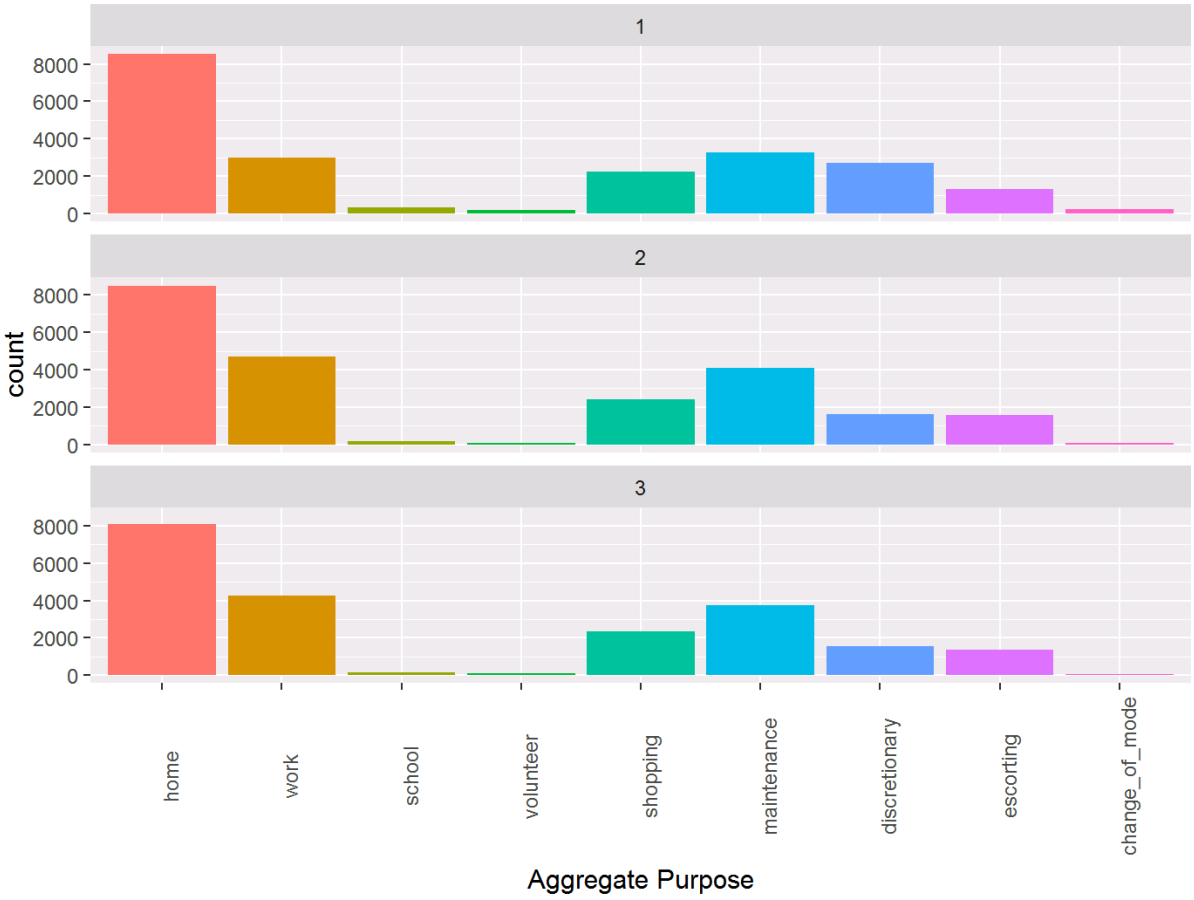


Figure 1 – Aggregate Purpose Distributions across Travel Days (1 = Observed, 2 and 3 = Imputed).

Appendix A: Computed Model Variables

Variables Used in the Decision Trees

Variable	Description or Formula
is_morning	arr_hr >= 6 AND arr_hr < 11
is_midday	arr_hr >= 11 AND arr_hr < 14
is_evening	arr_hr >= 14 AND arr_hr < 16
is_night	arr_hr >= 17 AND arr_hr < 23
mode	Travel mode for place.
next_mode	Travel mode for next place on the same day.
is_transit	mode IN (8,12,13,15)
is_next_transit	next_mode IN (8,12,13,15)
ploctype	Location type (1 = home, 2 = work, 3 = school, 4 = other)
is_other_member_location	Place's end location is either a work or school location for another household member.
longer_stop	Place is the longest non-home and non-tour anchoring stop of the day.
shopping_clusters	Place is within 250m of a shopping cluster.
maintenance_clusters	Place is within 250m of a maintenance cluster.
volunteer_clusters	Place is within 250m of a volunteer cluster.
discretionary_clusters	Place is within 250m of a discretionary cluster.
change_of_mode_clusters	Place is within 250m of a change of mode cluster.
escorting_points	Place is within 75m of a escorting PR place for this person.
work_points	Place is within 150m of a work PR place for this person.
longer_stop	Place has the longest activity duration for travel day.
subtour_stop	Place is a stop on a work subtour.

Appendix B: Decision Tree Specifications

Worker Tree

```
Call:
C5.0.formula(formula = formula, data = train, control = C5.0Control(CF = 0.25, minCases = 10))


C5.0 [Release 2.07 GPL Edition]          Fri Feb 26 15:19:41 2016
-------------------------------

Class specified by attribute `outcome'

Read 8785 cases (23 attributes) from undefined.data

Decision tree:

ploctype = 3:
:...actdur <= 44: escorting (383/32)
:   actdur > 44:
:   :...is_evening = 0: volunteer (34/24)
:       is_evening = 1: discretionary (16/5)
ploctype = 2:
:...is_other_member_location <= 0: work (1729/73)
:   is_other_member_location > 0:
:   :...actdur > 65: work (49/8)
:       actdur <= 65:
:       :...actdur <= 7: escorting (40/15)
:           actdur > 7:
:           :...age <= 47: escorting (13/7)
:               age > 47: work (11/4)
ploctype = 4:
:...shopping_clusters > 0:
    :...mode_changes > 0:
    :   :...change_of_mode_clusters > 0: change_of_mode (34/10)
    :   :   change_of_mode_clusters <= 0:
    :   :   :...aage in {-8,-7}: shopping (0)
    :   :       aage = 2: discretionary (1)
    :   :       aage = 1:
    :   :       :...maintenance_clusters <= 0: shopping (14/9)
    :   :           maintenance_clusters > 0:
    :   :           :...actdur <= 17: maintenance (12/3)
    :   :               actdur > 17: shopping (12/5)
    :   mode_changes <= 0:
    :   :...maintenance_clusters <= 0:
    :       :...discretionary_clusters <= 0:
    :       :   :...actdur <= 43: shopping (440/36)
    :       :   :   actdur > 43:
    :       :   :   :...is_morning = 0: shopping (74/16)
    :       :   :       is_morning = 1: work (13/4)
    :       :   discretionary_clusters > 0:
    :       :   :...actdur > 97: discretionary (20/8)
    :       :       actdur <= 97:
    :       :       :...sex in {-7,2}: shopping (91/32)
    :       :           sex = 1:
    :       :           :...is_midday = 1: discretionary (16/4)
    :       :               is_midday = 0:
    :       :               :...actdur <= 12: shopping (21/6)
    :       :                   actdur > 12: discretionary (30/13)
    :       maintenance_clusters > 0:
    :       :...discretionary_clusters <= 0:
    :           :...actdur > 14:
    :           :   :...actdur <= 182: shopping (329/86)
```

```
:           :   :   actdur > 182: work (12)
:           :   actdur <= 14:
:           :   :...actdur <= 5: maintenance (251/94)
:           :       actdur > 5:
:           :       :...is_morning = 1:
:           :           :...age <= 54: maintenance (22/9)
:           :           :   age > 54: shopping (11/3)
:           :           is_morning = 0:
:           :           :...is_midday = 0:
:           :               :...actdur <= 11: shopping (84/30)
:           :               :   actdur > 11: maintenance (27/9)
:           :               is_midday = 1:
:           :               :...age <= 37: maintenance (14/2)
:           :                   age > 37: shopping (28/8)
:           discretionary_clusters > 0:
:           :...actdur > 10:
:               :...actdur > 194: work (15/2)
:               :   actdur <= 194:
:               :   :...actdur > 49: discretionary (98/48)
:               :       actdur <= 49:
:               :       :...subtour_stop <= 0: shopping (215/108)
:               :           subtour_stop > 0: discretionary (25/12)
:               actdur <= 10:
:               :...actdur <= 1: escorting (25/16)
:                   actdur > 1:
:                   :...volunteer_clusters > 0: shopping (15/6)
:                       volunteer_clusters <= 0:
:                       :...is_evening = 0: maintenance (196/79)
:                           is_evening = 1:
:                           :...actdur <= 4: maintenance (33/13)
:                               actdur > 4: shopping (53/23)
shopping_clusters <= 0:
:...maintenance_clusters > 0:
    :...change_of_mode_clusters <= 0:
    :   :...discretionary_clusters <= 0:
    :   :   :...actdur <= 129: maintenance (1027/124)
    :   :   :   actdur > 129: work (81/23)
    :   :   discretionary_clusters > 0:
    :   :   :...actdur <= 23: maintenance (463/145)
    :   :       actdur > 23:
    :   :       :...actdur <= 225: discretionary (262/87)
    :   :           actdur > 225: work (19/7)
    :   change_of_mode_clusters > 0:
    :   :...mode_changes > 0: change_of_mode (36/10)
    :       mode_changes <= 0:
    :       :...aage in {-8,-7}: maintenance (0)
    :           aage = 2: change_of_mode (5/2)
    :           aage = 1:
    :           :...discretionary_clusters <= 0:
    :               :...actdur <= 16: maintenance (19/8)
    :               :   actdur > 16: work (21/9)
    :               discretionary_clusters > 0:
    :               :...actdur <= 12: maintenance (20/7)
    :                   actdur > 12: discretionary (27/12)
    maintenance_clusters <= 0:
    :...discretionary_clusters > 0:
        :...volunteer_clusters > 0:
        :   :...is_evening = 1: discretionary (27/10)
        :   :   is_evening = 0:
        :   :   :...actdur <= 21: escorting (10/4)
        :   :       actdur > 21: volunteer (10/6)
        :   volunteer_clusters <= 0:
        :   :...actdur > 22: discretionary (674/65)
        :       actdur <= 22:
        :       :...age > 49: discretionary (158/28)
        :           age <= 49:
        :           :...is_morning = 0: discretionary (189/72)
        :               is_morning = 1: escorting (34/12)
        discretionary_clusters <= 0:
```

```
:...volunteer_clusters > 0: volunteer (48/14)
    volunteer_clusters <= 0:
    :...change_of_mode_clusters > 0:
        :...mode_changes > 0: change_of_mode (45/4)
        :   mode_changes <= 0:
        :   :...actdur <= 20: escorting (32/14)
        :       actdur > 20: work (14/5)
        change_of_mode_clusters <= 0:
        :...actdur > 11: work (700/48)
            actdur <= 11:
            :...is_evening = 1:
                :...sex in {-7,2}: escorting (37/3)
                :   sex = 1:
                :   :...actdur <= 5: escorting (28/4)
                :       actdur > 5: work (10/2)
                is_evening = 0:
                :...actdur > 5: work (108/34)
                    actdur <= 5:
                    :...age > 55: work (63/23)
                        age <= 55:
                        :...is_midday = 0: escorting (89/26)
                            is_midday = 1:
                            :...age <= 38: work (10/2)
                                age > 38: escorting (13/3)
```

Evaluation on training data (8785 cases):

Decision Tree	
Size	Errors
70	1675(19.1%) <<

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	<-classified as
94	10	11	6		6	1	17	(a): class change_of_mode
	1179	22	118		95	8	39	(b): class discretionary
3	86	568	73		29	9	116	(c): class escorting
11	77	28	1591		161	2	43	(d): class maintenance
	4	19	5			3	8	(e): class school
3	39	4	154		1019		12	(f): class shopping
3	16	5	2		2	48	9	(g): class volunteer
6	132	47	135		75	21	2611	(h): class work

Attribute usage:

100.00% ploctype
77.93% actdur
74.10% shopping_clusters
73.71% maintenance_clusters
72.81% discretionary_clusters
36.45% change_of_mode_clusters
29.55% volunteer_clusters
27.89% mode_changes
20.97% is_other_member_location
8.39% is_evening
7.46% age
5.65% is_morning
3.78% is_midday
2.73% subtour_stop
2.65% sex
1.49% aage

Time: 0.2 secs

University Student Tree

```
formula <- tpurp ~ is_morning + is_midday + is_evening + is_night + age + aage + sex + lic + nolic +
mode + mode_changes + is_transit + is_next_transit + actdur + ploctype + is_other_member_location +
volunteer_clusters + maintenance_clusters

Call:
C5.0.formula(formula = formula, data = train, control = C5.0Control(CF = 0.25, minCases = 10))


C5.0 [Release 2.07 GPL Edition]          Mon Feb 15 14:03:47 2016
-------------------------------

Class specified by attribute `outcome'

Read 1082 cases (23 attributes) from undefined.data

Decision tree:

escorting_points > 0: escorting (150/19)
escorting_points <= 0:
:...ploctype = 3: school (109/14)
    ploctype = 2:
    :...is_other_member_location <= 0: work (121/5)
    :   is_other_member_location > 0: discretionary (17/11)
    ploctype = 4:
    :...shopping_clusters > 0:
        :...maintenance_clusters <= 0: shopping (75/8)
        :   maintenance_clusters > 0:
        :   :...actdur <= 28:
        :       :...actdur <= 7: maintenance (73/22)
        :       :   actdur > 7: shopping (56/22)
        :       actdur > 28:
        :       :...discretionary_clusters <= 0: shopping (15/2)
        :           discretionary_clusters > 0: discretionary (17/10)
        shopping_clusters <= 0:
        :...maintenance_clusters <= 0:
            :...discretionary_clusters > 0: discretionary (144/15)
            :   discretionary_clusters <= 0:
            :   :...change_of_mode_clusters > 0: change_of_mode (13/2)
            :       change_of_mode_clusters <= 0:
            :       :...aage in {-8,-7,1}: work (75/24)
            :           aage = 2: school (3/1)
            maintenance_clusters > 0:
            :...discretionary_clusters <= 0:
                :...actdur <= 134: maintenance (106/4)
                :   actdur > 134: work (10/5)
                discretionary_clusters > 0:
                :...change_of_mode_clusters > 0: discretionary (10/3)
                    change_of_mode_clusters <= 0:
                    :...actdur <= 52: maintenance (64/19)
                        actdur > 52: discretionary (24/12)
```

```
Evaluation on training data (1082 cases):

            Decision Tree
          ----------------
          Size      Errors

            18  198(18.3%)   <<


           (a)   (b)   (c)   (d)   (e)   (f)   (g)   (h)    <-classified as
          ----  ----  ----  ----  ----  ----  ----  ----
            11     3                 1                       (a): class change_of_mode
                 161    12    23    10     6           1    (b): class discretionary
                       131                                  (c): class escorting
                  12     1   198          23           4    (d): class maintenance
             1    10     2     1    97     1          25    (e): class school
                   5           17     1   114               (f): class shopping
                   1     1           1                 4    (g): class volunteer
             1    20     3     4     2     2         172    (h): class work


        Attribute usage:

        100.00% escorting_points
         86.14% ploctype
         63.31% maintenance_clusters
         63.31% shopping_clusters
         44.45% discretionary_clusters
         33.73% actdur
         17.47% change_of_mode_clusters
         12.75% is_other_member_location
          7.21% aage


Time: 0.0 secs
```

Non-Worker Tree

```
formula <- tpurp ~ is_morning + is_midday + is_evening + is_night + age + aage + sex + lic + nolic +
mode + mode_changes + is_transit + is_next_transit + actdur + ploctype + is_other_member_location +
volunteer_clusters + maintenance_clusters

Call:
C5.0.formula(formula = formula, data = train, control = C5.0Control(CF = 0.25, minCases = 10))


C5.0 [Release 2.07 GPL Edition]          Mon Feb 15 16:37:19 2016
-------------------------------

Class specified by attribute `outcome'

Read 2064 cases (19 attributes) from undefined.data

Decision tree:

is_transit > 0:
:...actdur <= 43: change_of_mode (47/4)
:   actdur > 43: maintenance (13/9)
is_transit <= 0:
:...ploctype = 3: escorting (161/22)
    ploctype = 2:
    :...actdur <= 30: escorting (19/4)
    :   actdur > 30: work (29/6)
    ploctype = 4:
    :...is_next_transit > 0: change_of_mode (24/5)
        is_next_transit <= 0:
        :...maintenance_clusters > 0:
            :...actdur <= 11: maintenance (523/151)
            :   actdur > 11:
            :   :...actdur > 67:
            :       :...is_evening = 0: maintenance (110/50)
            :       :   is_evening = 1: discretionary (17/5)
            :       actdur <= 67:
            :       :...is_evening = 1:
            :           :...mode = 1: maintenance (5/2)
            :           :   mode in {2,3,4,6,7,8,9,10,11,14,15,
            :           :   :        97}: shopping (42/25)
            :           :   mode = 5:
            :           :   :...actdur > 24: discretionary (10/4)
            :           :       actdur <= 24:
            :           :       :...aage in {-8,-7,1}: maintenance (14/6)
            :           :           aage = 2: shopping (7/2)
            :           is_evening = 0:
            :           :...sex = -7: discretionary (2/1)
            :               sex = 1: maintenance (84/43)
            :               sex = 2:
            :               :...aage in {-8,-7}: shopping (4/2)
            :                   aage = 2: maintenance (23/12)
            :                   aage = 1:
            :                   :...mode in {1,3,6,7,8,9,10,11,14,15,
            :                       :        97}: shopping (1)
            :                       mode = 2: maintenance (1)
            :                       mode = 5:
            :                       :...age <= 53: shopping (20/9)
            :                       :   age > 53: discretionary (14/5)
            :                       mode = 4:
            :                       :...actdur > 33: shopping (60/24)
            :                           actdur <= 33:
            :                           :...is_midday = 0:
            :                               :...age <= 48: shopping (26/11)
            :                               :   age > 48: maintenance (32/13)
            :                               is_midday = 1:
            :                               :...actdur <= 21: shopping (22/9)
```

```
:                                   actdur > 21: maintenance (13/6)
maintenance_clusters <= 0:
:...volunteer_clusters > 0:
    :...aage = -8: volunteer (0)
    :   aage = -7: discretionary (1)
    :   aage = 2: change_of_mode (3/2)
    :   aage = 1:
    :   :...actdur <= 13: escorting (14/6)
    :       actdur > 13: volunteer (20/7)
    volunteer_clusters <= 0:
    :...actdur > 71:
        :...is_evening = 1: discretionary (52/8)
        :   is_evening = 0:
        :   :...actdur <= 172: discretionary (75/18)
        :       actdur > 172: work (56/24)
        actdur <= 71:
        :...actdur <= 3:
            :...mode_changes <= 0: escorting (100/43)
            :   mode_changes > 0: discretionary (12/5)
            actdur > 3:
            :...actdur <= 16: shopping (167/93)
                actdur > 16:
                :...is_evening = 1:
                    :...actdur <= 25: shopping (14/3)
                    :   actdur > 25: discretionary (48/17)
                    is_evening = 0:
                    :...is_midday = 1: shopping (67/23)
                        is_midday = 0:
                        :...sex = -7: escorting (1)
                            sex = 1:
                            :...is_morning = 0: shopping (18/9)
                            :   is_morning = 1: discretionary (17/9)
                            sex = 2:
                            :...mode = 1: discretionary (3)
                                mode in {2,3,5,7,8,9,10,11,14,15,
                                :        97}: shopping (11/5)
                                mode = 6: work (1)
                                mode = 4:
                                :...is_morning = 0: shopping (35/20)
                                    is_morning = 1: [S1]

SubTree [S1]

age <= 37: shopping (14/5)
age > 37: discretionary (12/5)
```

```
Evaluation on training data (2064 cases):

            Decision Tree
          ----------------
          Size      Errors

            49  732(35.5%)   <<


           (a)   (b)   (c)   (d)   (e)   (f)   (g)   (h)    <-classified as
          ----  ----  ----  ----  ----  ----  ----  ----
            63     2     2     4           2     1          (a): class change_of_mode
             1   186    33    68         107     4    22    (b): class discretionary
             4    11   220    31          64           2    (c): class escorting
             1     6     7   526          48           1    (d): class maintenance
                   8     3     4           1           2    (e): class school
             3    33    13   159         268           3    (f): class shopping
                        9     6                13          (g): class volunteer
             2    17     8    20          18     2    56    (h): class work


        Attribute usage:

        100.00% is_transit
         97.09% ploctype
         90.84% actdur
         86.97% is_next_transit
         85.80% maintenance_clusters
         45.11% is_evening
         35.90% volunteer_clusters
         20.06% sex
         16.62% mode
         13.32% aage
         13.18% is_midday
          5.72% age
          5.43% mode_changes
          4.65% is_morning


Time: 0.1 secs
```

Retiree Tree

```
formula <- tpurp ~ is_morning + is_midday + is_evening + is_night + lic + nolic + mode + mode_changes +
is_transit + is_next_transit + actdur + ploctype + is_other_member_location + volunteer_clusters +
maintenance_clusters + shopping_clusters + discretionary_clusters + change_of_mode_clusters +
escorting_points + work_points

Call:
C5.0.formula(formula = formula, data = train, control = C5.0Control(CF = 0.25, minCases = 10))


C5.0 [Release 2.07 GPL Edition]         Mon Feb 15 14:04:23 2016
-------------------------------

Class specified by attribute `outcome'

Read 2709 cases (21 attributes) from undefined.data

Decision tree:

escorting_points > 0: escorting (241/31)
escorting_points <= 0:
:...shopping_clusters <= 0:
    :...volunteer_clusters > 0:
    :   :...maintenance_clusters <= 0: volunteer (80/11)
    :   :   maintenance_clusters > 0:
    :   :   :...ploctype = 3: maintenance (0)
    :   :       ploctype = 2: work (1)
    :   :       ploctype = 4:
    :   :       :...actdur <= 108: maintenance (35/14)
    :   :           actdur > 108: volunteer (12/2)
    :   volunteer_clusters <= 0:
    :   :...work_points > 0: work (21)
    :       work_points <= 0:
    :       :...discretionary_clusters <= 0: maintenance (458)
    :           discretionary_clusters > 0:
    :           :...maintenance_clusters <= 0: discretionary (456/2)
    :               maintenance_clusters > 0:
    :               :...actdur <= 17: maintenance (125/26)
    :                   actdur > 17: discretionary (161/43)
    shopping_clusters > 0:
    :...discretionary_clusters <= 0:
        :...maintenance_clusters <= 0: shopping (310/1)
        :   maintenance_clusters > 0:
        :   :...actdur <= 9: maintenance (151/58)
        :       actdur > 9: shopping (248/69)
        discretionary_clusters > 0:
        :...actdur <= 14:
            :...maintenance_clusters <= 0: shopping (22/4)
            :   maintenance_clusters > 0: maintenance (145/72)
            actdur > 14:
            :...mode_changes > 0: discretionary (14/5)
                mode_changes <= 0:
                :...maintenance_clusters <= 0:
                    :...actdur <= 51: shopping (46/17)
                    :   actdur > 51: discretionary (28/6)
                    maintenance_clusters > 0:
                    :...actdur <= 73: shopping (116/50)
                        actdur > 73: discretionary (39/15)
```

```
Evaluation on training data (2709 cases):

            Decision Tree
          ----------------
          Size      Errors

            20  426(15.7%)   <<


           (a)   (b)   (c)   (d)   (e)   (f)   (g)   (h)    <-classified as
          ----  ----  ----  ----  ----  ----  ----  ----
                                                             (a): class change_of_mode
                 627    17    43          56    12          (b): class discretionary
                       210                                   (c): class escorting
                  52     9   744          82     1          (d): class maintenance
                   3                                         (e): class school
                  15     1   117         601                (f): class shopping
                         4    10           3    79          (g): class volunteer
                   1                                   22    (h): class work


        Attribute usage:

        100.00% escorting_points
         91.10% shopping_clusters
         85.60% discretionary_clusters
         72.91% maintenance_clusters
         49.80% volunteer_clusters
         45.07% work_points
         42.16% actdur
          8.97% mode_changes
          1.77% ploctype


Time: 0.0 secs
```

Driving Age Student Tree

```
formula <- tpurp ~ is_morning + is_midday + is_evening + is_night + mode_changes + is_transit +
is_next_transit + actdur + ploctype + is_other_member_location + maintenance_clusters +
shopping_clusters + discretionary_clusters + longer_stop

Call:
C5.0.formula(formula = formula, data = train, control = C5.0Control(CF = 0.25, minCases = 10))


C5.0 [Release 2.07 GPL Edition]          Mon Feb 15 14:04:50 2016
-------------------------------

Class specified by attribute `outcome'

Read 349 cases (16 attributes) from undefined.data

Decision tree:

actdur > 85:
:...is_evening = 0: school (137/19)
:   is_evening = 1: discretionary (27/7)
actdur <= 85:
:...maintenance_clusters > 0: maintenance (73/36)
    maintenance_clusters <= 0:
    :...shopping_clusters > 0: shopping (11/2)
        shopping_clusters <= 0:
        :...discretionary_clusters <= 0: escorting (61/17)
            discretionary_clusters > 0:
            :...ploctype = 3: school (11/5)
                ploctype in {2,4}: discretionary (29/4)


Evaluation on training data (349 cases):

            Decision Tree
          ----------------
          Size      Errors

             7   90(25.8%)   <<


           (a)   (b)   (c)   (d)   (e)   (f)   (g)   (h)    <-classified as
          ----  ----  ----  ----  ----  ----  ----  ----
                   3     4     3     1                      (a): class change_of_mode
                  45           13    10     1               (b): class discretionary
                   1    44     1     5                      (c): class escorting
                              37     1                      (d): class maintenance
                   6     8     2   124                      (e): class school
                              16           9               (f): class shopping
                         1           1                      (g): class volunteer
                   1     4     1     6     1               (h): class work


        Attribute usage:

        100.00% actdur
         53.01% maintenance_clusters
         46.99% is_evening
         32.09% shopping_clusters
         28.94% discretionary_clusters
         11.46% ploctype


Time: 0.0 secs
```

Appendix XX

Pilot Report

(Appendices not included)

MI Travel Counts III

Final Pilot Survey Report

Authors

Jesse Casas, Project Director
Jeremy Wilhelm, Project Manager
Shawn McCloskey, Project Manager

April 2015

Prepared for:
The Michigan Department of Transportation
Van Wagner Building
425 West Ottawa Street
Lansing, MI 48933
(517) 373-2090

and

The Southeast Michigan Council of Governments
1001 Woodward Ave., Suite 1400
Detroit, MI 48226
(313) 961-4266

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, MD 20850-3129
(301) 251-1500

This page is intentionally left blank.

Table of Contents

Tables

Figures

Equations

1. Introduction

This report provides a summary of the design and execution of the pretest and pilot surveys that were used to prepare for the MI Travel Counts III survey conducted in 2015.

Work began on the design and development of the materials, instruments, software, websites, and public outreach efforts in October of 2014, with the goal of delivering a simulation of the survey experience for a staff pretest in January of 2015. Feedback from the staff pretest was then incorporated before a pilot survey was conducted in February of 2015, in which 11,290 households were randomly selected from the general public and invited to participate. The focus of the pilot survey was not to achieve a preset number of completed households; rather, the pilot was undertaken as a means of testing the methods and systems developed in collaboration among staff from Westat, the Michigan Department of Transportation (MDOT), and the Southeast Michigan Council of Governments (SEMCOG).

Specifically, these exercises were intended to identify weaknesses or shortcomings of the data collection tools, to seek out opportunities for enhancement of all processes, and to reveal any unanticipated procedural or data quality shortcomings. Therefore, the sample selection process did not use a proportionally representative approach but rather an even allocation of sampled households throughout the state. Households with targeted socio-demographic characteristics were also not included in the sampling effort for the pilot. Participation rates for the pilot were useful for assessing assumptions about recruitment and retrieval rates, and for adjusting the number of invited households for the spring data collection phase, but were not intended to evaluate the performance of the sample in specific two-way aggregate cells (e.g., household size by number of workers), or for comparison against American Community Survey (ACS) benchmarks.

2. Pilot Survey Approach

2.1 Staff Pretest

A staff pretest was conducted with the aim of testing the survey systems before inviting the public to take part in the pilot survey. For this design, a total of 38 staff members from MDOT and SEMCOG were recruited to take part in the pretest. Staff members were a mix of key managers from each agency along with additional staff who were either completely unfamiliar, or at least less familiar, with the procedures or purpose of the survey. The goal for the pretest was to simulate an actual participant's experience as closely as possible. However, due to time constraints, there were several key differences between the "real" and "simulated" experiences, including:

- Mailed letters of invitation were not sent or presented;
- No helpdesk or computer-assisted telephone interviewing (CATI) mode was available, either for support or completion of the survey;
- Automated text and email reminders were not provided;
- Personalized travel logs and instructions were not mailed; and
- Some tutorials and aids normally available to participants were not yet available for pre-testers.

Westat provided 59 samples and personal identification numbers (PINs) for use by the staff, and assigned at least one PIN to each person, while some participants received multiple PINs. Selected staff members were then sent an email with instructions on how to participate in the pretest. The email included a link to the pretest web address; instructions for entering the PIN; a generic travel

log along with an example log; and instructions about the timeline for recruiting into the pretest, reporting travel, and providing feedback.

The pretest was conducted using the web survey mode, in part because the majority of participants normally elect this mode. It was deemed prudent to have the feedback received based on the most prevalent mode used. Further, the underlying structure, logic, and presentation of the two modes (web and CATI) was similar enough that observations for one mode were deemed to be pertinent to the other.

Westat received consolidated comments from MDOT and SEMCOG staff for each stage of the survey (recruitment and retrieval). Changes were implemented in time for the pilot survey wherever possible, but postponed until the main survey in select situations.[1]

2.2 Pilot Survey

For the full pilot test, letters of invitation were sent to members of the public, who were provided with a complete survey experience. This included the receipt of the invitation letter as well as two reminder postcards; access to the complete, public-facing website; access to helpdesk technical support via a toll-free hotline; the choice to complete the survey by web or telephone; receipt (per stated preference) of a full battery of phone, email, and text message reminders; and a monetary incentive. A subsample of 10% of households was invited to take part in a Global Positioning System (GPS) and Prompted Recall (PR) survey (GPS+PR), in which eligible household members would carry a GPS device, return it to Westat, and later view their processed data to confirm travel details. The purpose of the pilot GPS+PR effort was to assess the effectiveness of the recruitment, participant instructions, and the survey instrument and post-processes.

The pilot survey was conducted from January 2015 through March 2015 with sample selection occurring in early January. Data collection commenced with the mailing of 11,290 invitations on January 26, 2015, with travel dates assigned from February 9, 2015, through March 6, 2015. Recruitment data were delivered on March 9, 2015. The collection of travel details ended on March 15, 2015, and retrieval data were delivered on March 30, 2015. The pilot effort included a notable public awareness campaign.

2.3 Public Awareness

For the pilot survey, public awareness activities reflected the majority of the efforts planned for the main survey. One item not leveraged in the pilot was the series of informational videos produced by MDOT staff. These videos have been posted in conjunction with the launch of the main survey.

The public awareness methods used in the pilot included press releases by MDOT and SEMCOG on January 27, 2015, to coincide with the mailing of invitations on January 26, 2015. Concurrently, letters were mailed to local governments so that public officials would be knowledgeable about the survey. Officials were asked to encourage their invited constituents to take part in the survey. All public awareness materials can be viewed in Appendix D.

Social media sites were leveraged to raise exposure, including the creation of accounts on Facebook and Twitter. These sites provided links to the public website (and vice versa). Any media coverage

[1] Documentation of the comments and changes were provided via email to MDOT and SEMCOG project managers on 1/27/2015 for recruitment and 2/10/2015 for retrieval. See Appendix C.

was linked from the "News about the Survey" section of the public site. The goal with these efforts was to provide a sense of legitimacy for prospective participants.

The public site also provided a means for participants or prospective participants to send questions or feedback to the survey team. A member of the team responded to each comment, and in cases where the message included a complaint or observation about the transportation system in Michigan, these were forwarded to MDOT and SEMCOG staff for review and response as appropriate.

3. Pilot Results

3.1 Sampling and Demographic Outcomes

The following section provides results and analysis of the performance of the pilot survey including discussion about response rates, the performance of key variables, and a summary of item non-response. The section concludes with a caveat on the limitations of interpreting the survey results from a sample composition perspective.

3.1.1 Survey Mode by Survey Stage – All Households

The modes offered to households participating in the pilot included both web and CATI surveys. There was no incentive offered to participate by one mode over the other, which means that the mode selected by the participant likely represents his or her unbiased preference. Table 1 shows the total and percent of survey mode for each survey stage. The columns headed "CATI In" show the count and percent of cases that completed by an incoming call to the helpdesk, whereas the columns headed "CATI Out" show the count and percent of cases that were called by the helpdesk. The recruitment stage did not utilize active outbound calling efforts, so all calls either were initiated by the participant calling the helpdesk, or by the helpdesk responding to an email or voicemail message.

Table 1. Share of households by survey mode and survey stage

	WEB		CATI						Total Households
	Count	%	CATI In	CATI In %	CATI Out	CATI Out %	All CATI	All CATI %	
Recruit	748	89%	94	11%	0	0%	94	11%	842
Retrieval	386	69%	48	9%	126	23%	174	31%	560

3.1.2 Response Rates – All Households

Target and actual response rates are summarized in Table 2, which shows that the actual rates for each category were better than anticipated. This resulted in a higher than expected number of completed pilot surveys, which in turn led to changes in the mailing strategy for the first phase of the main survey data collection.

Table 2. Sample performance rates; targeted versus actual

Rates	Targeted	Actual
Recruitment Rate[2]	5%	7.5%
Participation Rate[3]	62%	66.4%

[2] The percent of households that joined the survey (based on the number of invitations mailed)
[3] The percent of recruited households that reported travel

Completion Rate[4]	3.1%	5%
Completed Surveys	350	560

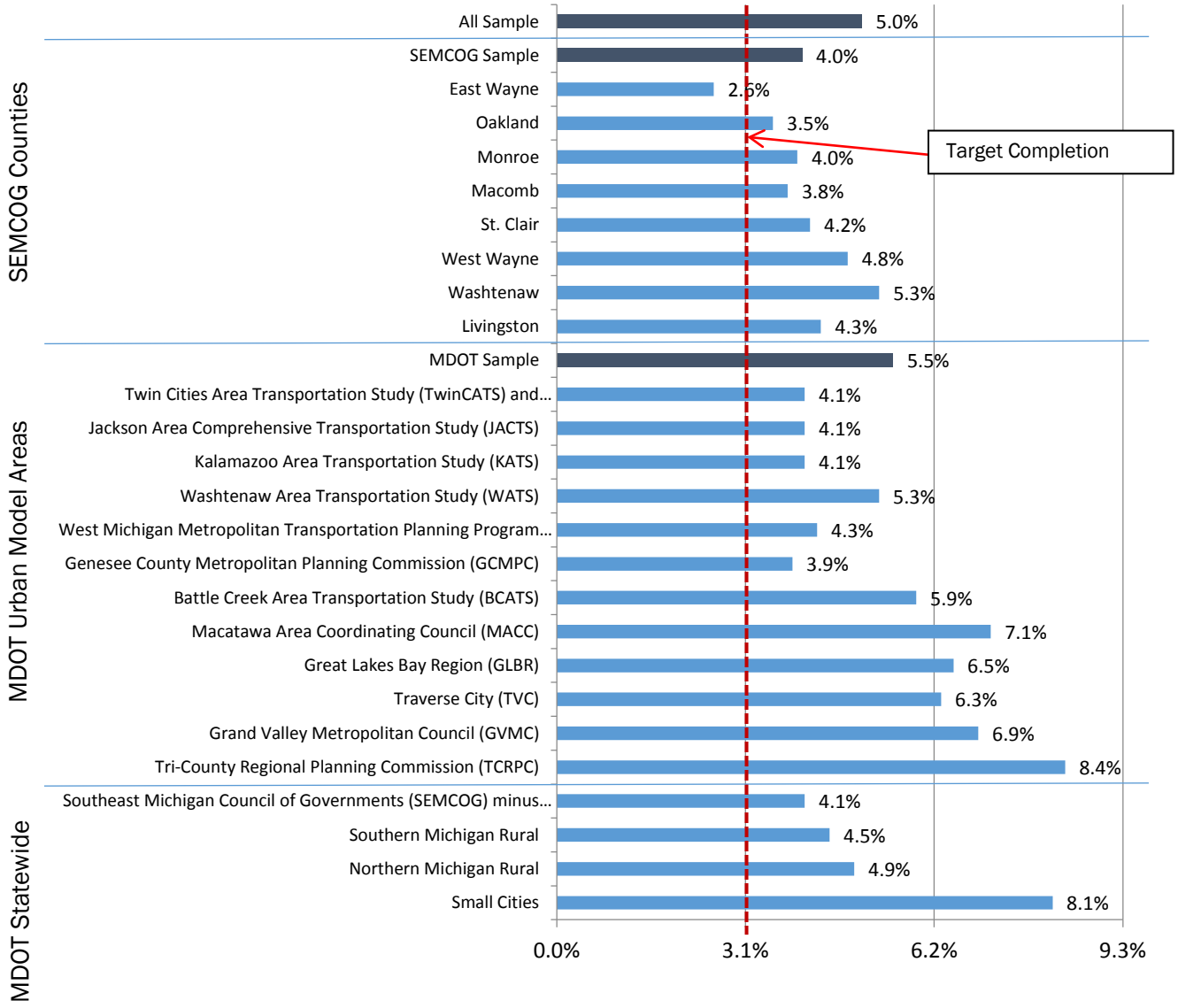
Table 3 and Figure 1 show the results aggregated to the model area or county with subtotals for the MDOT statewide model, MDOT urban model areas, and SEMCOG counties, in addition to the overall survey performance. Overall participation rates exceeded targeted rates, as did the rates for each model area or county in the sampling plan, with the one exception being East Wayne County, as seen in Table 3.

[4] The product of the recruitment rate and the completion rate

Table 3. Participation rates by model area/county

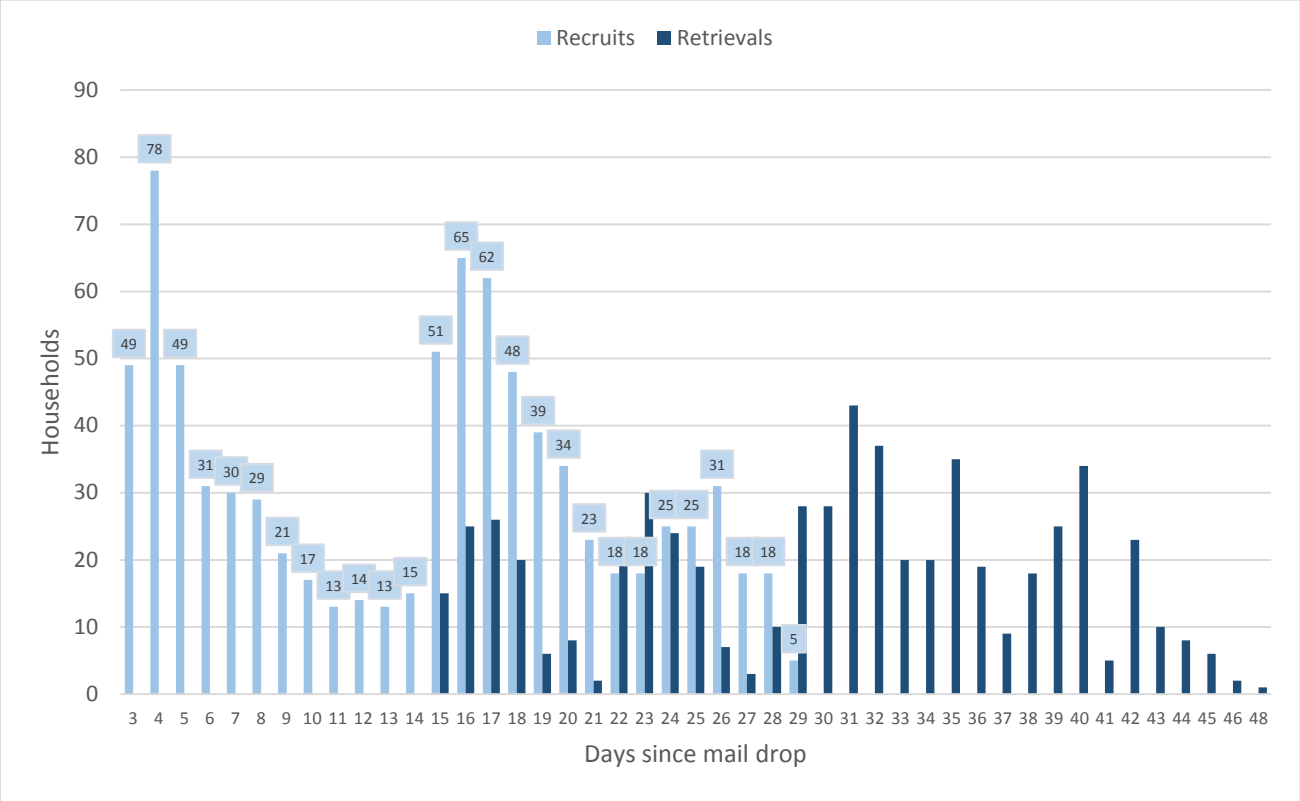
Sample Area	Recruitment Rate	Participation Rate	Completion Rate
Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County (WATS)	4.3%	90.5%	3.9%
Southern Michigan Rural	5.5%	77.8%	4.3%
Northern Michigan Rural	10.0%	44.9%	4.5%
Small cities	10.8%	73.6%	7.9%
Statewide Model Total	**7.6%**	**67.3%**	**5.1%**
Urban Model Areas			
Grand Valley Metropolitan Council (GVMC)	10.2%	68.0%	6.9%
Tri-County Regional Planning Commission (TCRPC)	12.8%	65.1%	8.4%
Genesee County Metropolitan Planning Commission (GCMPC)	7.5%	48.6%	3.7%
Great Lakes Bay Region (GLBR)	9.6%	68.1%	6.5%
Washtenaw Area Transportation Study (WATS)	6.9%	73.5%	5.1%
Kalamazoo Area Transportation Study (KATS)	6.9%	58.8%	4.1%
West Michigan Metropolitan Transportation Planning Program (WestPlan)	7.3%	61.1%	4.5%
Jackson Area Comprehensive Transportation Study (JACTS)	6.5%	62.5%	4.1%
Twin Cities Area Transportation Study (TwinCATS) and Niles/Buchanan/Cass Area Transportation Study (NATS)	5.5%	74.1%	4.1%
Macatawa Area Coordinating Council (MACC)	9.2%	77.8%	7.1%
Battle Creek Area Transportation Study (BCATS)	8.4%	70.7%	5.9%
Traverse City (TVC)	10.0%	63.3%	6.3%
Urban Model Areas Total	**8.4%**	**66.1%**	**5.5%**
Combined MDOT Total	**8.2%**	**66.4%**	**5.4%**
SEMCOG Counties			
East Wayne	4.1%	62.5%	2.6%
West Wayne	6.4%	75.0%	4.8%
Oakland	4.7%	75.9%	3.5%
Macomb	5.2%	70.0%	3.6%
Washtenaw	6.9%	73.5%	5.1%
Monroe	5.1%	76.9%	4.0%
St. Clair	5.9%	70.0%	4.2%
Livingston	7.9%	55.0%	4.3%
SEMCOG Total	**5.7%**	**69.6%**	**4.0%**
Grand Total	**7.5%**	**66.4%**	**5.0%**
Targets	**5.0%**	**62.0%**	**3.1%**

Figure 1. Completion rates by sample area or county



The numbers presented in Figure 2 were used in Equation 1 to derive an average number of days to respond for recruited households by taking a weighted average; i.e., the average of: the product of the count of households and the number of days since the mail drop for each day of recruitment, divided by the count of recruited households (839). The result of 14.28 provides an average number of days that the overall sample took to respond to the initial invitation.

Figure 2. Recruited and retrieved households by day



Equation 1. Weighted average of recruits by count of days from mail-out

$$\bar{x} = \frac{(R^1 \times D^1) + (R^2 \times D^2) + \cdots (R^{29} \times D^{29})}{N}$$

Where:

R = Count of Recruits on day D

D = Number of days since the mail drop

N = Total number of recruited households

3.1.3 Days to Complete – Log-Only Households

A common indicator of survey quality is the volume of households that report travel within 1 week of the assigned travel date. A rate of 65% or more of households reporting within one week of assigned travel dates, and 75% or more within two weeks, indicates a responsive sample. Figure 3 shows results for MTC III/STC 15 along with two recent, similar surveys. The comparison shows that pilot survey responses came at a higher rate in the early days following the travel date and that the majority of households reported within one week. For this survey, the metric applies exclusively to log-only households because there is a built-in delay with GPS+PR households (i.e., there is a time lag between GPS returns, processing, and prompted recall).

Figure 3. Comparison of number of days to report by survey

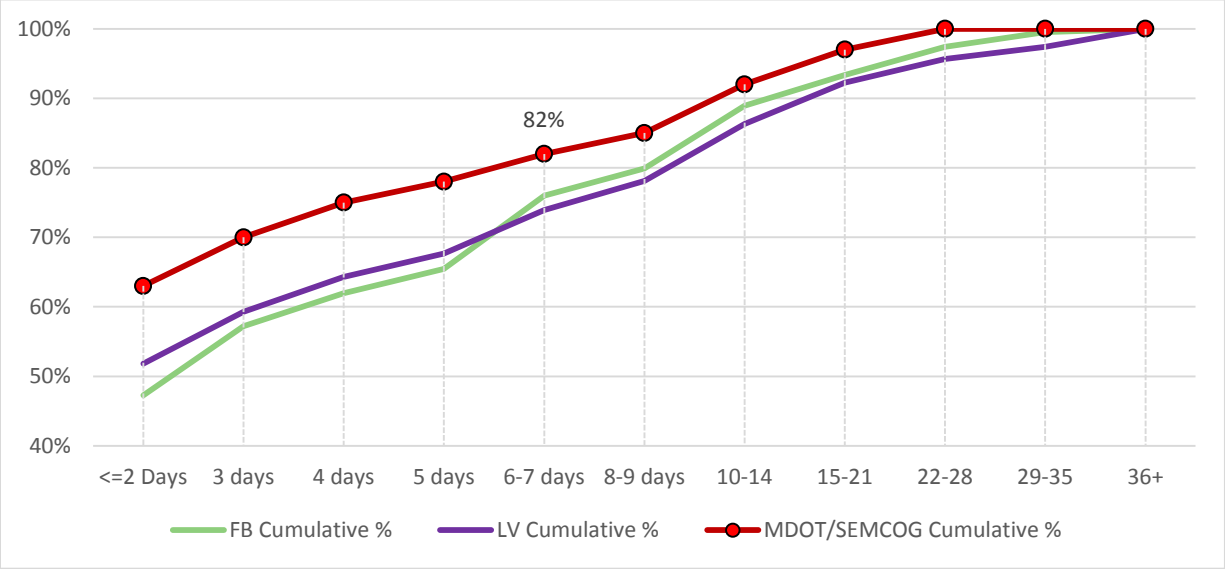


Table 4 presents the results for log-only households and shows that 92% of these households responded within 2 weeks of their travel date, 82% of households responded within 1 week, and 63% completed within 3 days. Only 8.2% of households took more than 2 weeks to respond.

Table 4. Days to complete from first travel date – Log Only

Number of Days	Count	% of Total	Cumulative	Cum. %
Fewer than 3 days	325	63.1%	325	63%
3 days	34	6.6%	359	70%
4 days	28	5.4%	387	75%
5 days	17	3.3%	404	78%
6-7 days	20	3.9%	424	82%
8-9 days	14	2.7%	438	85%
10-14 days	35	6.8%	473	92%
15-21 days	29	5.6%	502	97%
22-28 days	11	2.1%	513	100%
29-35 days	2	0.4%	515	100%

3.1.4 Income Responses – All Households

The next section deals with responses to income and shows responses for MDOT and SEMCOG separately, beginning with MDOT results. Income is a common variable with high instances of non-response (i.e., many households answer with a "Don't know" or "Refused to answer" response). To mitigate this, the survey utilized a second income question with broader categories. The second question was asked only for households that completed retrieval and did not respond to the initial income question in the recruitment step. Achieving non-response rates of 10% or lower was considered a reasonable target.

Income for responding MDOT households saw a non-response rate of 13.2%. Table 5 shows that the households that completed the travel reporting step had 52 non-responses. The follow-up question elicited a response from 34 of those households and reduced overall income non-response from 12.1% to 4.2%, as shown in Table 6.

Table 5. Reported income for MDOT sample

	Recruited Households		Retrieved Households	
Income	**Actual**	**% of Total**	**Actual**	**% of Total**
Less than $15,000	76	11.8%	45	10.5%
$15,000 - $ 24,999	62	9.6%	45	10.5%
$25,000 - $34,999	64	9.9%	44	10.3%
$35,000 - $49,999	84	13.0%	49	11.4%
$50,000 - $74,999	121	18.8%	91	21.3%
$75,000 - $99,999	59	9.1%	39	9.1%
$100,000 - $124,999	45	7.0%	28	6.5%
$125,000 - $149,999	20	3.1%	13	3.0%
$150,000 or more	29	4.5%	22	5.1%
Refused	75	11.6%	46	10.7%
Don't know	10	1.6%	6	1.4%
Totals	645	100.0%	428	100.0%

Table 6. Income follow-up for MDOT sample

Income	**Actual**	**% of Total**
Less than $25,000	99	23.1%
$25,000 - $49,999	99	23.1%
$50,000 - $74,999	98	22.9%
$75,000 or more	114	26.6%
Refused	17	4.0%
Don't know	1	0.2%
Totals	428	100.0%

The SEMCOG sample in the recruitment step showed a rate of 15.8% non-response to the income question, as seen in Table 7. Households that completed the final travel reporting step had a 15.3% non-response rate to the initial question. The follow-up question elicited a response from 11 of 27 non-response households, with the effect of reducing the overall non-response to 9.1%, as seen in Table 8.

Table 7. Reported income for SEMCOG sample

	Recruited Households		Retrieved Households	
Income	**Actual**	**% of Total**	**Actual**	**% of Total**
Less than $15,000	17	6.7%	14	8.0%
$15,000 - $ 24,999	20	7.9%	12	6.8%
$25,000 - $34,999	22	8.7%	16	9.1%
$35,000 - $49,999	29	11.5%	17	9.7%
$50,000 - $74,999	37	14.6%	27	15.3%
$75,000 - $99,999	28	11.1%	16	9.1%
$100,000 - $124,999	29	11.5%	22	12.5%
$125,000 - $149,999	7	2.8%	5	2.8%
$150,000 or more	24	9.5%	20	11.4%
Refused	40	15.8%	27	15.3%
Don't know	0	0.0%	0	0.0%
Totals	253	100%	176	100%

Table 8. Income follow-up for SEMCOG sample

Income	**Actual**	**% of Total**
Less than $25,000	28	15.9%
$25,000 - $49,999	35	19.9%
$50,000 - $74,999	29	16.5%
$75,000+	68	38.6%
Refused	16	9.1%
Don't know	0	0.0%
Totals	176	100.0%

3.1.5 Trip Rates – All Households

The following section provides counts of households aggregated by the total count of trips and, separately, the average trips per person. In this analysis, the trips are unlinked, meaning a trip is counted for any movement from one location to another, regardless of purpose.

For the MDOT sample, the distribution of households by the number of trips appears to be reasonable based on historical trip rates and the patterns that Westat has observed in prior survey efforts, with a peak of household trips coming between 4 and 6, followed by a long tail of observations with higher trip counts, as seen in Table 9. Notably, a fairly high number of households had more than 18 trips, accounting for 14.7% of all reporting households in the MDOT sample.

Table 9. Trips per household for MDOT sample

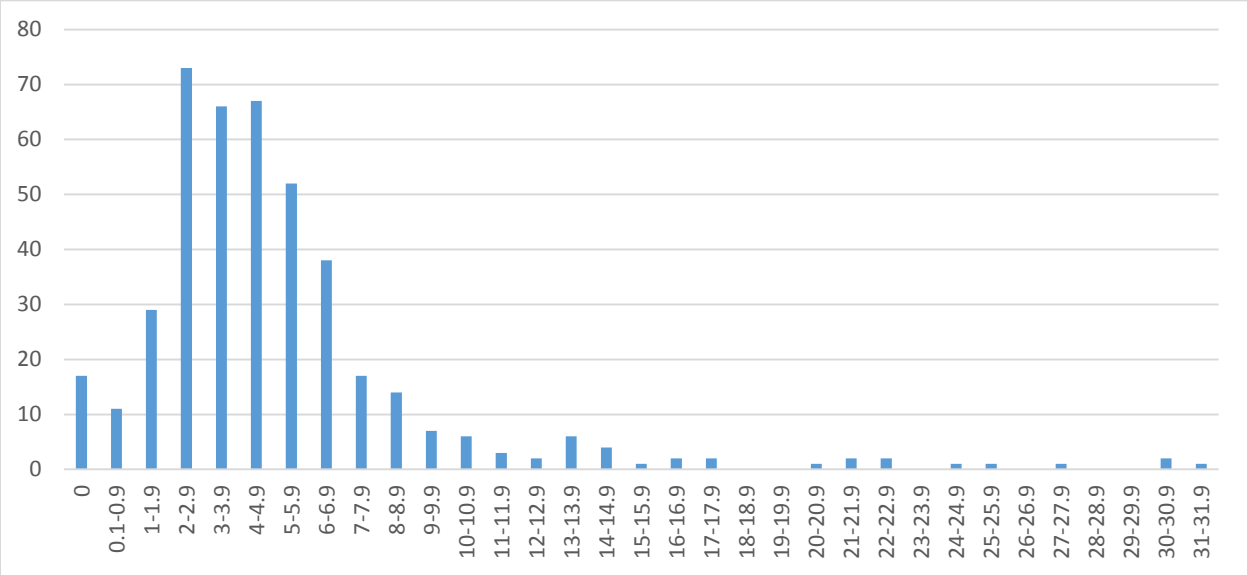
Trips	Households	Percent
0	17	4.0%
1	1	0.2%
2	39	9.1%
3	16	3.7%
4	48	11.2%
5	38	8.9%
6	42	9.8%
7	21	4.9%
8	30	7.0%
9	17	4.0%
10	22	5.1%
11	13	3.0%
12	18	4.2%
13	9	2.1%
14	7	1.6%
15	7	1.6%
16	10	2.3%
17	10	2.3%
18+	63	14.7%
Total	428	100.0%

Similarly, the data in Table 10 (and Figure 4) show a peak of between 2 and 5.9 trips per person, with a long tail as trip numbers increase. The majority (86.4%) of households average 8 or fewer trips per person in a day, and more than 60.3% of households average between 2 and 6 trips per person.

Table 10. Household count by trips per person for MDOT sample

Trips Per Person	Households	Percent
0	17	4.0%
0.1 to 0.9	11	2.6%
1 to 1.9	29	6.8%
2 to 2.9	73	17.1%
3 to 3.9	66	15.4%
4 to 4.9	67	15.7%
5 to 5.9	52	12.1%
6 to 6.9	38	8.9%
7 to 7.9	17	4.0%
8 to 8.9	14	3.3%
9 to 9.9	7	1.6%
10 to 10.9	6	1.4%
11 to 11.9	3	0.7%
12 to 12.9	2	0.5%
13+	26	6.1%
Total	428	100.0%

Figure 4. Chart of household count by trips per person for MDOT sample



For the SEMCOG sample, the distribution of households by the number of trips appears to be slightly more dispersed than with the MDOT sample, with a peak of trips landing between 4 and 8 household trips, followed by a long tail of observations with higher trip counts, as seen in Table 11.

Table 11. Trips per household for SEMCOG sample

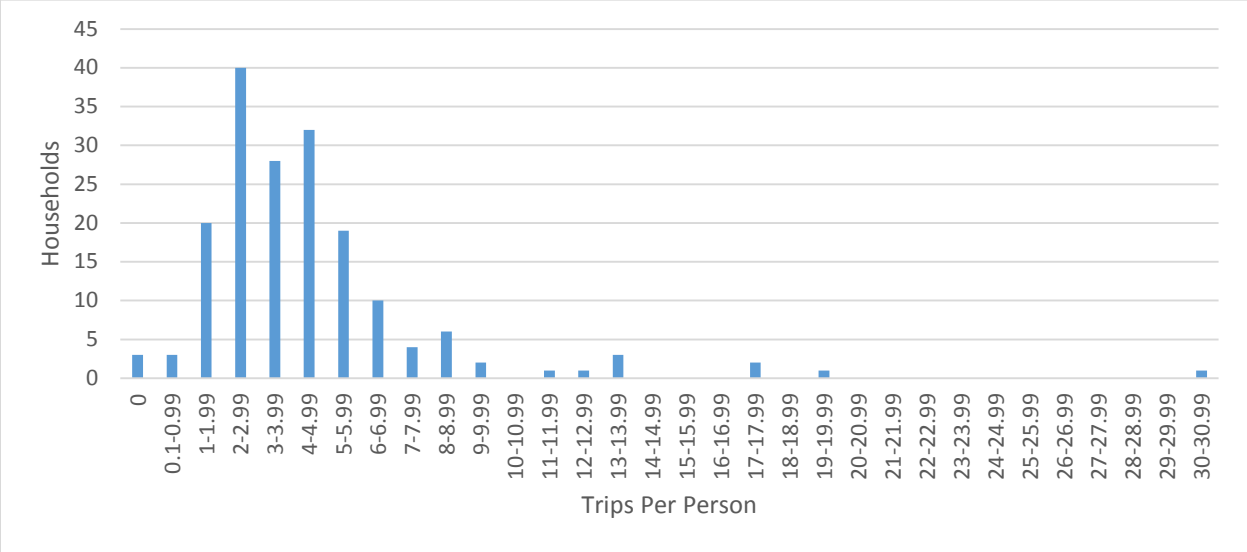
Trips	Households	Percent
0	3	1.7%
1	1	0.6%
2	12	6.8%
3	11	6.3%
4	26	14.8%
5	19	10.8%
6	12	6.8%
7	11	6.3%
8	20	11.4%
9	6	3.4%
10	7	4.0%
11	10	5.7%
12	5	2.8%
13	5	2.8%
14	4	2.3%
15	1	0.6%
16	3	1.7%
17	3	1.7%
18+	17	9.7%
Total	**176**	**100.0%**

Similarly, the data in Table 12 (and Figure 5) show a peak of between 2 and 4 trips per person, with a long tail as trip numbers increase. The majority (90.3%) of households average 8 or fewer trips per person in a day, and 67.6% of households average between 2 and 6 trips per person.

Table 12. Household count by trips per person for SEMCOG sample

Trips Per Person	Households	Percent
0	3	1.7%
0.1 to .9	3	1.7%
1 to 1.9	20	11.4%
2 to 2.9	40	22.7%
3 to 3.9	28	15.9%
4 to 4.9	32	18.2%
5 to 5.9	19	10.8%
6 to 6.9	10	5.7%
7 to 7.9	4	2.3%
8 or more	17	9.7%
Total	**176**	**100**

Figure 5. Chart of household count by trips per person for SEMCOG sample



3.1.6 Age Responses – All Households

There is some discussion in the survey community about whether or not web surveys can provide representative results due to limitations regarding technology for some segments of the population. Commonly cited populations tend to be those who are low-income and those who are elderly. To consider these concerns, Westat analyzed responses by age category to indicate whether or not the pilot survey could inform the question. It is notable that though the pilot survey was heavily web-administered, there was still an over-representation of elderly respondents in both MDOT and SEMCOG samples as seen in Figures 7 and 9.

Figure 6 shows the count of persons for each of the age categories presented in the age follow-up question for MDOT households. The people who responded to the first age category have been rolled into their respective categories. The figure does not contain a count of "Refused" responses because there were none collected for the pilot.

Figure 6. Count of persons by age category for MDOT sample

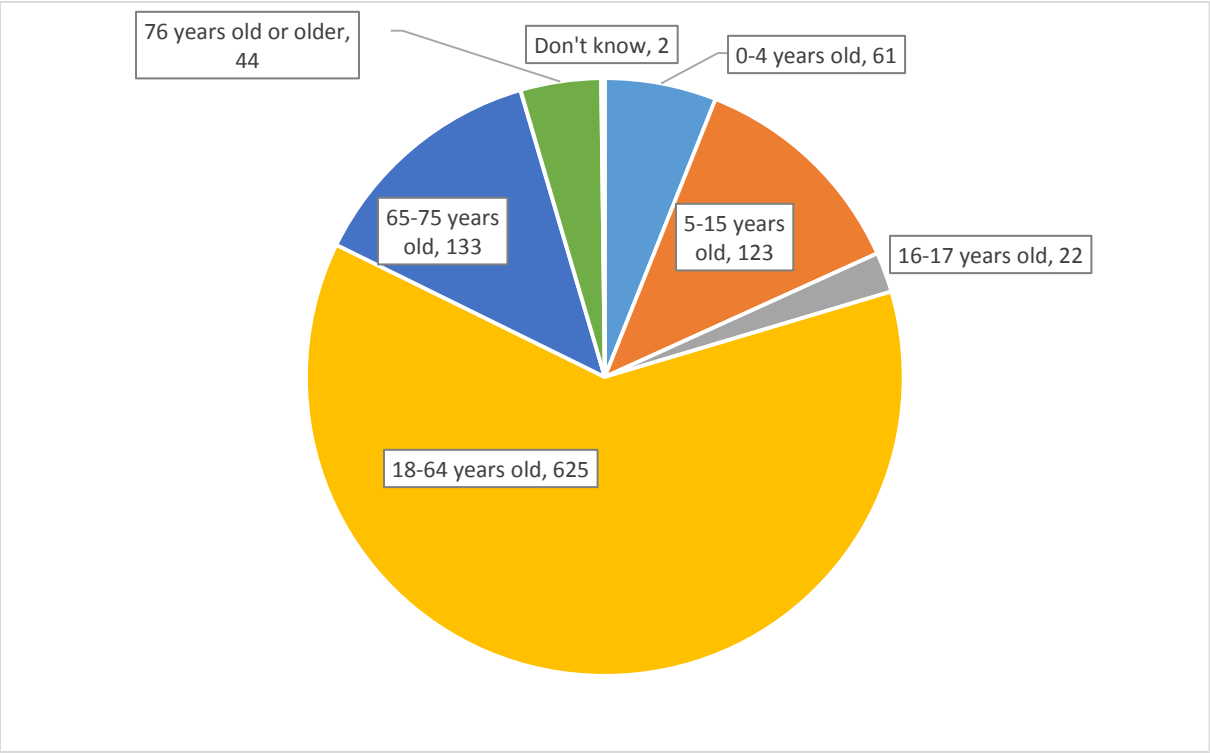
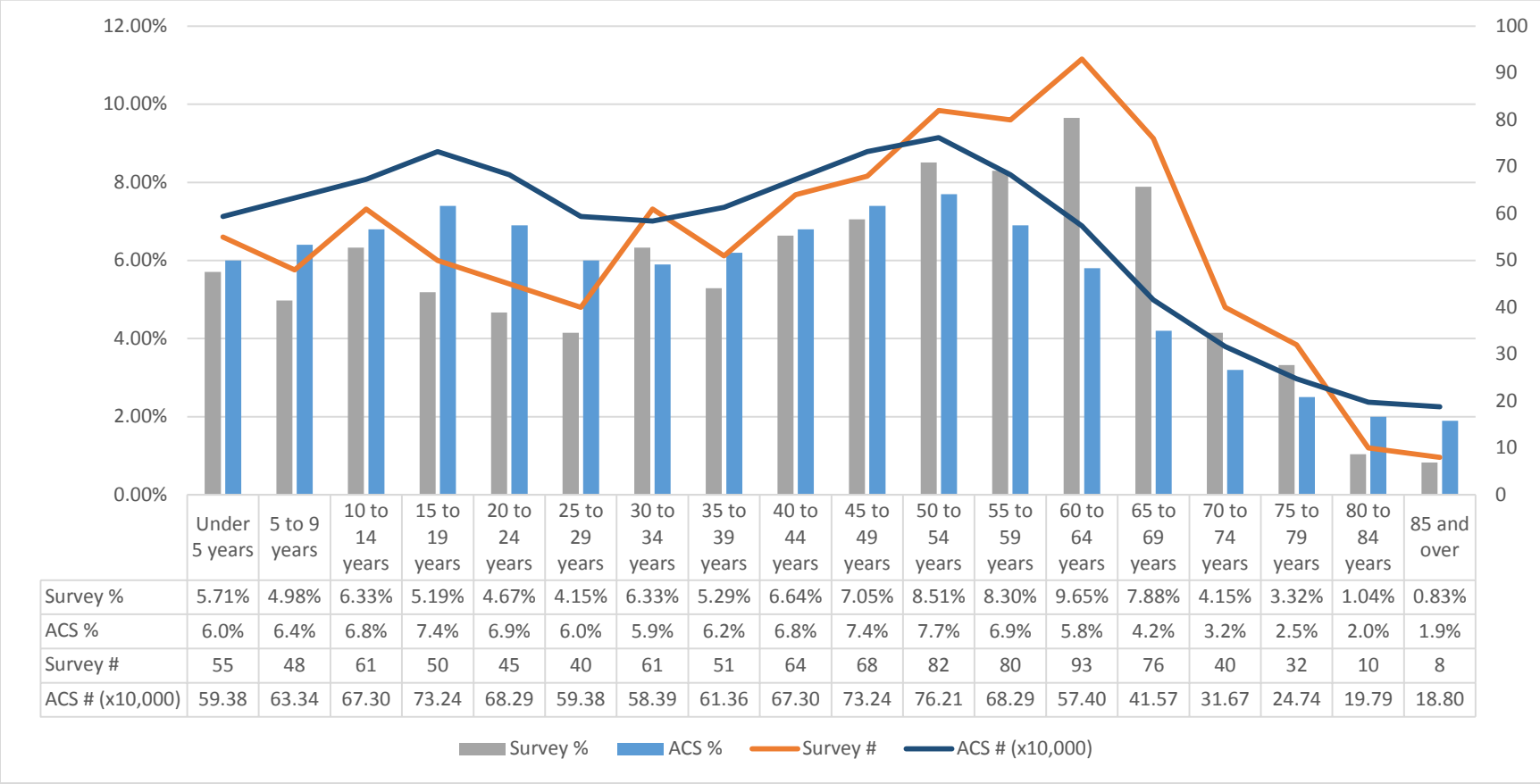


Figure 7 contains the responses to the AGE question for MDOT households aggregated to ACS age bins and includes 2008-2012 ACS benchmarks for comparison. Responses from survey participants in the 60 to 64 and 65 to 69 age categories were notably higher than ACS benchmarks. The limitation of the pilot sample design and the relatively small sample size should be taken into account while reviewing these results. With that in mind, Westat does not suggest changes to the main survey based on the distribution of age in the pilot sample.

Figure 7. Age categories compared to 2008-2012 ACS for state of Michigan



	Under 5 years	5 to 9 years	10 to 14 years	15 to 19 years	20 to 24 years	25 to 29 years	30 to 34 years	35 to 39 years	40 to 44 years	45 to 49 years	50 to 54 years	55 to 59 years	60 to 64 years	65 to 69 years	70 to 74 years	75 to 79 years	80 to 84 years	85 and over
Survey %	5.71%	4.98%	6.33%	5.19%	4.67%	4.15%	6.33%	5.29%	6.64%	7.05%	8.51%	8.30%	9.65%	7.88%	4.15%	3.32%	1.04%	0.83%
ACS %	6.0%	6.4%	6.8%	7.4%	6.9%	6.0%	5.9%	6.2%	6.8%	7.4%	7.7%	6.9%	5.8%	4.2%	3.2%	2.5%	2.0%	1.9%
Survey #	55	48	61	50	45	40	61	51	64	68	82	80	93	76	40	32	10	8
ACS # (x10,000)	59.38	63.34	67.30	73.24	68.29	59.38	58.39	61.36	67.30	73.24	76.21	68.29	57.40	41.57	31.67	24.74	19.79	18.80

Figure 8 shows the count of persons for each of the age categories presented in the age followup question for SEMCOG households. The people who responded to the first age category have been aggregated into their respective categories. The figure does not contain a count of “Refused” responses because there were none collected.

Figure 8. Count of persons by age category for SEMCOG sample

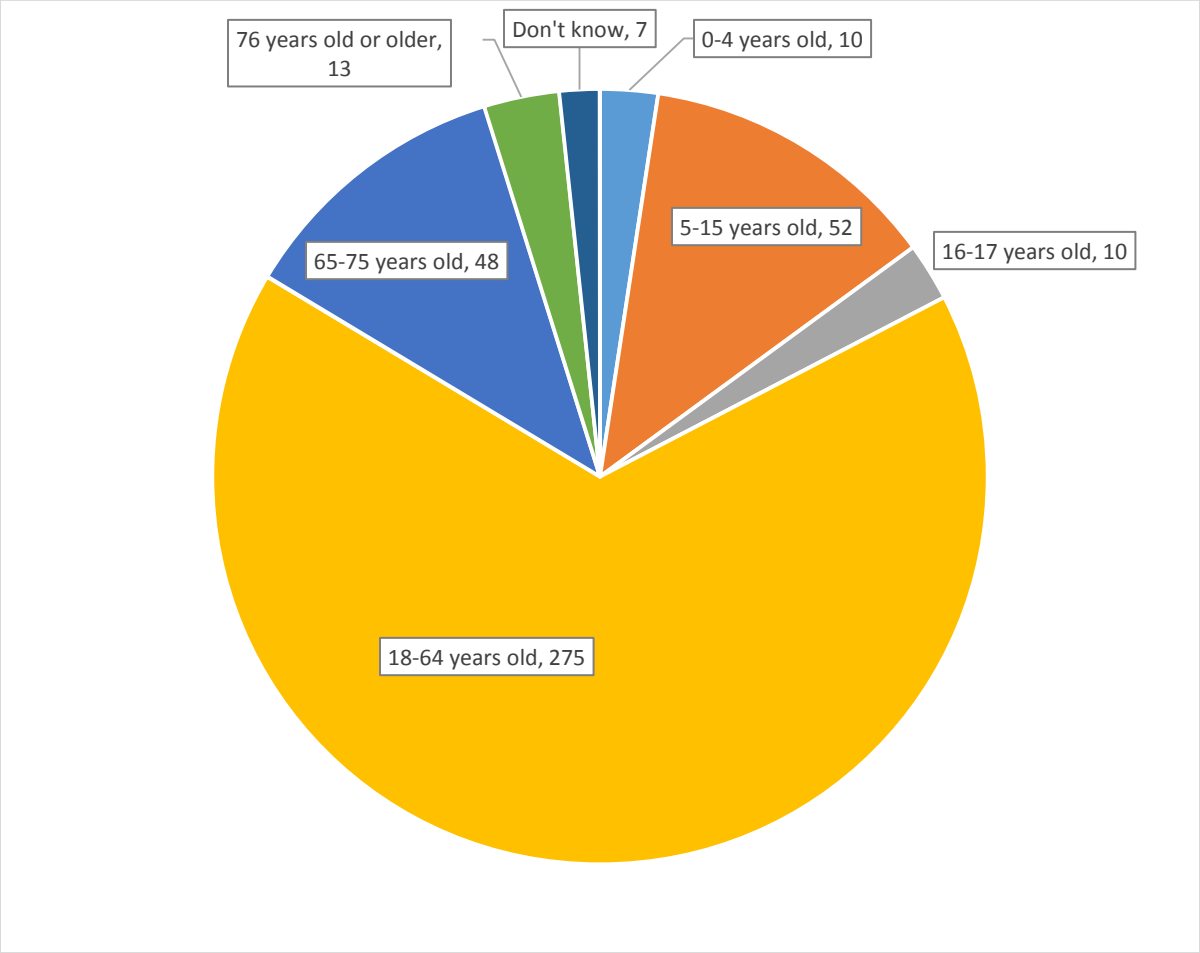
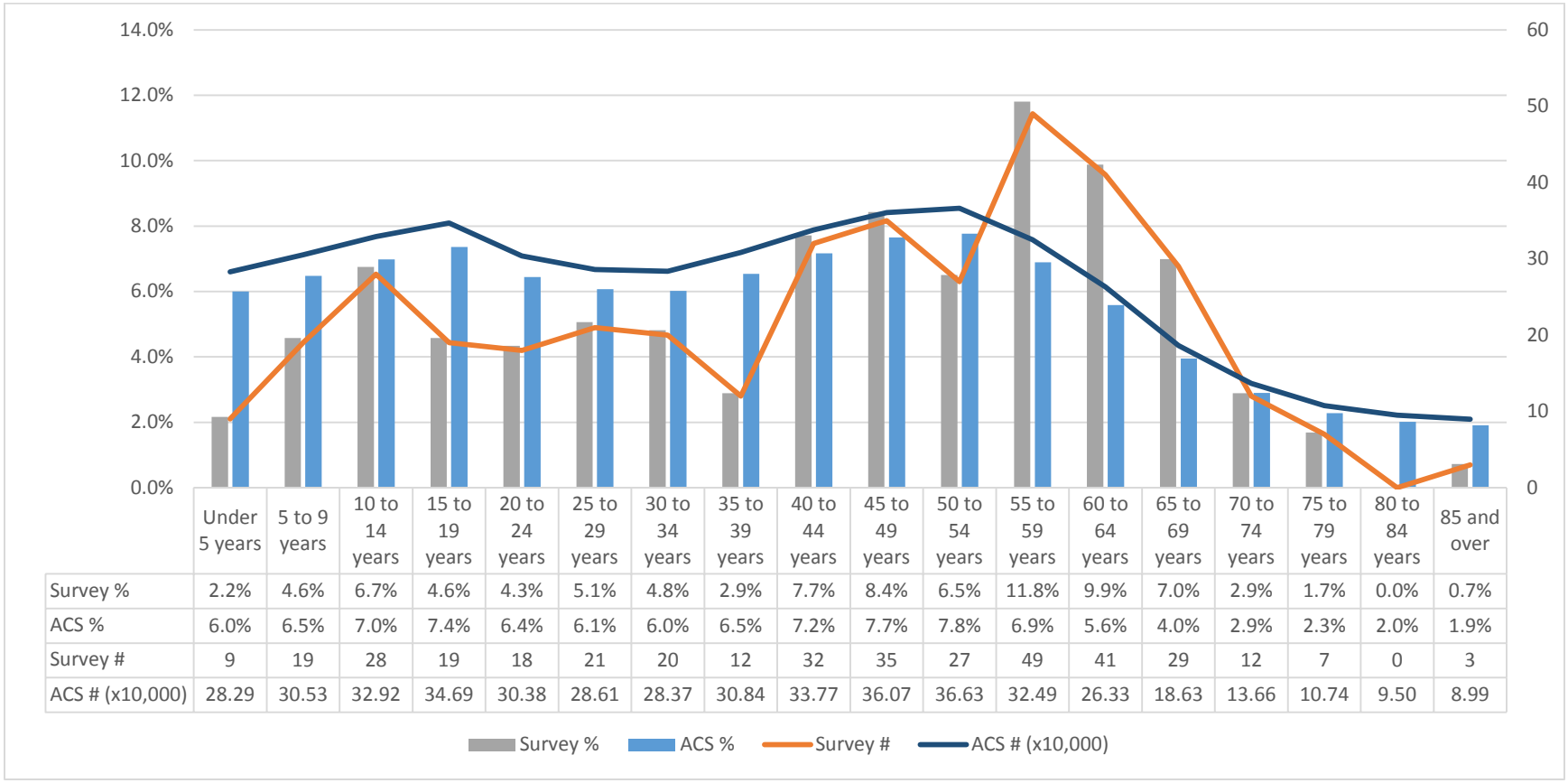


Figure 9 contains the responses to the AGE question for SEMCOG households aggregated to ACS age bins and includes 2008-2012 ACS benchmarks for comparison. Responses from survey participants in the 50 to 54, 55 to 59, and 60 to 64 age categories were notably higher than ACS benchmarks. The limitation of the pilot sample design and the relatively small sample size should be taken into account while reviewing these results. With that in mind, Westat does not suggest changes to the main survey based on the distribution of age in the pilot sample.

Figure 9. Age categories compared to 2008-2012 ACS for SEMCOG counties



	Under 5 years	5 to 9 years	10 to 14 years	15 to 19 years	20 to 24 years	25 to 29 years	30 to 34 years	35 to 39 years	40 to 44 years	45 to 49 years	50 to 54 years	55 to 59 years	60 to 64 years	65 to 69 years	70 to 74 years	75 to 79 years	80 to 84 years	85 and over
Survey %	2.2%	4.6%	6.7%	4.6%	4.3%	5.1%	4.8%	2.9%	7.7%	8.4%	6.5%	11.8%	9.9%	7.0%	2.9%	1.7%	0.0%	0.7%
ACS %	6.0%	6.5%	7.0%	7.4%	6.4%	6.1%	6.0%	6.5%	7.2%	7.7%	7.8%	6.9%	5.6%	4.0%	2.9%	2.3%	2.0%	1.9%
Survey #	9	19	28	19	18	21	20	12	32	35	27	49	41	29	12	7	0	3
ACS # (x10,000)	28.29	30.53	32.92	34.69	30.38	28.61	28.37	30.84	33.77	36.07	36.63	32.49	26.33	18.63	13.66	10.74	9.50	8.99

3.1.7 Item Non-Response

A review of item non-response is another tool available for assessing the effectiveness of survey questions. Table 13 provides a list of variables other than income, with percentages of non-response. The high rates in "Future Surveys" are the result of the fact that the variable was not implemented immediately at the start of the pilot survey. "Telecommuting offered" and "Home ownership status" also have relatively high levels of non-response. This is attributable to a lack of understanding the question in the first instance; not all people work in environments where this is an option, so the question can be confusing. Home ownership can be considered a personal question, and it is also possible that participants cannot connect the purpose of the question with the purpose of the survey. No changes are suggested for these variables.

Table 13. Item non-response

Variable	MDOT	SEMCOG
Age	0.2%	1.2%
Gender	0.6%	2.0%
Licensed	0.5%	1.5%
Employment status	0.1%	1.5%
Employer description	2.8%	6.5%
Industry	0.8%	2.5%
Telecommuting offered	7.4%	7.9%
School level among students	1.0%	5.7%
Home ownership status	4.5%	5.7%
Future surveys	11.2%	6.0%
Trip purpose	0.5%	1.4%

3.1.8 Limitations in Interpreting Sample Composition

The results of the pilot survey were intended to inform the project team about the viability of the survey design and systems, to identify systematic errors in the instrument or other systems, and to reveal opportunities for enhancement or improvement to those systems. Estimated response rates were used to inform the main survey but were balanced with a conservative adjustment to expectations.[5]

The goal was not to test the sampling plan or to confirm that the sampling cell targets could be achieved. To that end, the sample drawn for the pilot was not designed to yield a proportionally representative result. Therefore, overall response rates may or may not translate to the main survey. The results of various socio-demographic distributions will likely not be similar to the main survey results, in which the sample plan is designed to achieve specific targets. In the main survey, Westat will leverage the observed responses from the spring data collection to inform possible adjustments to the sample selection during the fall data collection period.

[5] More is covered on the topic of the mail-out plan and the changes made based on observed response rates in the "Updates to the Main Survey" section of this report.

3.2 Pilot Results - GPS and Prompted Recall Households

The purpose of the subsample of GPS+PR households for the main survey will be to provide trip rate correction factors, which can inform the rate of trip underreporting for the main survey sample. GPS+PR provides a way for participants to confirm GPS trips and details, or to add trips not captured by GPS, while viewing their GPS data and their travel logs together. This approach allows for a travel day collection that leverages the strengths of the GPS technology and the strengths of the travel log methodology while mitigating the weaknesses of each.

For the pilot survey, the GPS+PR approach was tested to ascertain the effectiveness of the recruitment, deployment, and retrieval of devices to households across the state of Michigan to confirm that the program provided a usable experience for participants to report prompted recall data.

To accomplish the GPS+PR data collection for the pilot, 40% of the 11,290 sampled households were flagged as "selected" for the GPS+PR subsample. A limit of 58 recruits (to yield 35 completes) was set and the numbers were distributed evenly across the pilot survey travel dates. Households selected for this effort were then invited to the GPS sample during the recruitment survey if they had at least one person in the household between the ages of 16 and 75. Households were not assigned to the GPS sample if the available dates to participate in the GPS survey were filled beyond 21 days from the then-current date.

Households that accepted the invitation into the GPS+PR survey were told that they would earn $25 per person assigned a GPS device, assuming that all members of the households completed all steps of the survey. (The per-person incentive came in lieu of the $20 household incentive. This point was clarified based on feedback received from the pretest.)

GPS devices were used for 3 consecutive days with the starting day falling at some point on a Monday through Thursday. Devices were returned on the fourth (or in the case of Thursday households, the fifth) day via FedEx. Data from returned devices was uploaded, processed, and released to participants for their review. Participants were informed that their data was ready for review by email, text message, or telephone call, based on their preference collected in the recruitment stage. Finally, responses in the confirmation step for completed GPS+PR households were processed with data checks and either cleared for delivery, sent to research, or closed out as unresolvable.

Table 14 shows the distribution of households by number of days between the travel date and the completion of travel confirmation for GPS+PR households only. Most GPS+PR households returned their devices to Westat between 6 and 11 days after the first of their three assigned travel dates and completed the travel confirmation step between 7 and 12 days after that date. Households in this cohort account for 71.1% of all GPS completes.

Table 14. Days to complete from first travel date – GPS households

Number of Days	Count	% of Total	Cumulative	Cum. %
Day 6	1	2.2%	1	2%
Day 7	3	6.7%	4	9%
Day 8	11	24.4%	15	33%
Day 9	4	8.9%	19	42%
Day 10	2	4.4%	21	47%
Day 11	8	17.8%	29	64%
Day 12	3	6.7%	32	71%
Day 13	2	4.4%	34	76%
Day 14	2	4.4%	36	80%
Day 15	1	2.2%	37	82%
Day 16	2	4.4%	39	87%
Day 17	2	4.4%	41	91%
Day 20	1	2.2%	42	93%
Day 22	1	2.2%	43	96%
Day 26	1	2.2%	44	98%
Day 29	1	2.2%	45	100%

The GPS+PR sub-sample in the pilot was successful, though the small number of households recruited limits any conclusions about how the main survey GPS effort will perform. Table 15 reports GPS progress for each sample area or county and contains columns for the following:

- Recruited – households that joined the survey and have been or will be sent GPS equipment (all households for the pilot fall under the "have been sent" category at the time of this report)
- Returned – households that were sent GPS equipment and have returned it to Westat
- Released – households that have had GPS data processed and are cleared to confirm travel details
- Completed – households that have confirmed all details and completed the retrieval survey
- Retrieval Rate – the percentage of recruited households that returned GPS equipment and completed the travel detail confirmation step

The rate of response in the retrieval step also outperformed expectations, with an expected overall retrieval rate of 60% and an actual overall rate of 65%.

Table 15. GPS participation by model area or county

Sample Area	Recruited	Returned	Released	Completed	Retrieval Rate
Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County (WATS)	3	3	3	2	67%
Southern Michigan Rural	1	1	1	1	100%
Northern Michigan Rural	5	4	4	3	60%
Small cities	7	7	7	7	100%
Statewide Model Total	**16**	**15**	**15**	**13**	**81%**
Grand Valley Metropolitan Council (GVMC)	3	2	2	2	67%
Tri-County Regional Planning Commission (TCRPC)	6	5	5	3	50%
Genesee County Metropolitan Planning Commission (GCMPC)	5	4	4	3	60%
Great Lakes Bay Region (GLBR)	3	3	3	3	100%
Washtenaw Area Transportation Study (WATS)	3	3	3	1	33%
Kalamazoo Area Transportation Study (KATS)	5	4	4	1	20%
West Michigan Metropolitan Transportation Planning Program (WestPlan)	2	2	2	1	50%
Jackson Area Comprehensive Transportation Study (JACTS)	1	1	1	0	-
Twin Cities Area Transportation Study (TwinCATS) and Niles/Buchanan/Cass Area Transportation Study (NATS)	4	3	3	2	50%
Macatawa Area Coordinating Council (MACC)	5	5	5	5	100%
Battle Creek Area Transportation Study (BCATS)	2	2	2	1	50%
Traverse City (TVC)	2	2	2	2	100%
Urban Model Areas Total	**41**	**36**	**36**	**24**	**59%**
Combined MDOT Total	**57**	**51**	**51**	**37**	**65%**
East Wayne	0	0	0	0	-
West Wayne	4	4	4	3	75%
Oakland	0	0	0	0	-
Macomb	2	2	2	1	50%
Washtenaw	3	3	3	1	33%
Monroe	4	4	4	4	100%
St. Clair	3	2	2	1	33%
Livingston	1	0	0	0	-
SEMCOG Total	**17**	**15**	**15**	**10**	**59%**
Grand Total	**68**	**60**	**60**	**44**	**65%**

Table 16 shows a comparison of person trip rates aggregated to the modeling area/county for each geographic area in the sample. The count of the households in the GPS areas shows that the highest count of households for any area was seven (with 17 people). These small numbers render the results difficult to interpret because the presence of outliers can dramatically impact trip rates. While the overall rates for log and GPS households are very close, the log rates have far lower variance (0.24 for log households versus 2.20 for GPS households), which is likely the result of the low accuracy (and perhaps a higher standard error) of the small GPS sample size. Westat recommends revisiting this question with the larger spring data set.

Table 16. Trip rate comparison between log-only and GPS+PR persons by sample area/county

Sample Area/County	Log-only Households	Log-only Persons	Log-only Person Trip Rates	GPS Households	GPS Persons	GPS Person Trip Rates
Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County (WATS)	19	47	3.79	2	6	5.39
Southern Michigan Rural	21	46	3.31	1	3	5.67
Northern Michigan Rural	22	46	4.11	3	6	6.33
Small cities	39	63	3.98	7	17	3.31
Grand Valley Metropolitan Council (GVMC)	34	72	4.41	2	8	5.38
Tri-County Regional Planning Commission (TCRPC)	41	72	4.43	3	7	5.43
Genesee County Metropolitan Planning Commission (GCMPC)	18	23	4.59	3	6	5.67
Great Lakes Bay Region (GLBR)	32	60	4.40	3	5	4.00
Washtenaw Area Transportation Study (WATS)	25	7	4.00	1	3	6.67
Kalamazoo Area Transportation Study (KATS)	20	46	4.20	1	2	5.50
West Michigan Metropolitan Transportation Planning Program (WestPlan)	22	53	4.29	1	2	3.00
Jackson Area Comprehensive Transportation Study (JACTS)	20	50	4.93	0	-	-
Twin Cities Area Transportation Study (TwinCATS) and Niles/Buchanan/Cass Area Transportation Study (NATS)	20	47	3.73	2	4	5.00
Macatawa Area Coordinating Council (MACC)	35	73	4.13	5	9	4.33
Battle Creek Area Transportation Study (BCATS)	29	68	4.29	1	2	2.50
Traverse City (TVC)	31	60	4.13	2	7	2.80
East Wayne	15	28	5.15	0	-	-
West Wayne	30	72	4.20	3	8	3.25
Oakland	22	45	3.82	0	-	-
Macomb	21	43	3.37	1	1	2.00
Washtenaw	25	47	4.34	1	3	6.67
Monroe	20	33	4.38	4	11	4.50
St. Clair	21	47	3.25	1	4	2.00
Livingston	22	63	3.16	0	-	-
Total	**560**	**1211**	**4.12**	**44**	**114**	**4.29**

4. Summary of Updates to the Main Survey

Improvements identified for data collection and survey processes are detailed in Table 17. The table contains a brief description of the change, a category for what type of element was changed, a category for the change type, and a brief note of explanation about the change (when needed). More detailed descriptions of the changes can be found in Appendix C. Note that materials, instruments or questions, or data elements not included in Table 17 were not changed.

The "Survey Element" categories include:

1. Materials – updates made to respondent materials, including letters, postcards, logs, etc.
2. Website –updates made to the public website
3. Data –updates made to the codebook or data elements
4. Instruments – updates made to the survey instruments
5. Process – updates made to a process or protocol

The "Change Type" categories include:

1. Wording – a minor update to wording to improve clarity
2. Sequencing – a change in question order to improve clarity
3. Logic – a change in logic to fix an unintentional skip
4. Enhancement – a general improvement to the survey experience
5. Data type – a change to the data type or formatting of a variable
6. Addition – an introduction of a new element or variable
7. Other – a change of some other type (specified in the notes)

Table 17. Summary of changes made to main survey based on pretest and pilot feedback

Item #	Change Description	Survey Element	Change Type	Note	Appendix C Location
1	Order of long trip questions	Instruments	Sequencing	Reordered for better clarity	1.1.0
2	GPS incentive explanation	Instruments	Wording	Reworded to make clear that GPS incentive is in lieu of standard incentive	1.1.1
3	Person roster headers	Instruments	Enhancement	Addition of summary headers for person roster to clarify who the question is directed to	1.1.2
4	Filtering city or county centroids	Instruments	Enhancement	Removed centroid results to improve quality of geocodes	1.1.3
5	Update to base home geocodes	Data	Enhancement	Re-geocoded all home locations not matched to a point or street address level	1.1.4
6	Mail-out strategy	Process	Enhancement		1.1.5
7	Quantity of invitations	Process	Enhancement		1.1.6
8	GPS Place name collection	Instruments	Enhancement		1.1.7
9	Long distance mode update	Instruments	Wording		1.1.8
10	Arabic letter update	Materials	Wording		1.1.9
11	Education question	Instruments	Logic		1.1.10
12	Household vehicle used	Instruments	Logic		1.1.11
13	Future surveys question added	Instruments	Addition		1.1.12
14	GPS device use and return instructions	Materials	Wording		1.1.13
15	GPS equipment usage log	Materials	Wording		1.1.14
16	Log cover letter	Materials	Wording		1.1.15
17	Add REGION variable to deliverable	Data	Addition		3.10.1
18	Error in LOCATION file character fields	Data	Data type		3.10.2
19	Update real numbers to integers	Data	Data type		3.10.3
20	Travel dates after original last pilot travel date	Process	Other		3.12.1
21	Mismatched sample region assignment	Data	Other	Recoding required	3.12.2

Item #	Change Description	Survey Element	Change Type	Note	Appendix C Location
22	Update REGION 17-23 to REGION 1	Data	Other	Recoding of cases from SEMCOG to MDOT	3.12.3
23	Missing HHVEH in HOUSEHOLD file	Data	Addition		3.12.4
24	Geocoding issue; placed at centroid of county	Data	Enhancement	This will be addressed by removing centroids per item 1.1.3	3.12.5
25	HHCHILD calculating incorrectly	Data	Logic		3.12.6
26	Fill AAGE based on answers in AGE?	Data	Addition	Westat will add a new age variable that combines AGE and AAGE	3.12.7
27	Eight records with blank geocode types	Data	Enhancement	Will flag cases for review in post processing	3.20.1
28	map_drag for a home location incorrect?	Data	None	map_drag should happen only if the original MSG/NAVTEQ geocode is below a precision of sample_mp or sample_ma	3.20.2
29	Person who is 1 has EMPTYPE answer and WPLACE answer	Data	None		3.20.3
30	Non workers with responses to work questions	Data	None		3.20.4
31	Blank State records	Data	Other	Fixed	3.20.5
32	AGE18 not populating correctly	Data	Other	Fixed	3.20.6
33	Change appropriate skips to Null?	Data	Other	Fixed	1.1
34	RETMODE of 3?	Data	Other	Fixed	1.2
35	hhtrips=pudotrips, shoptrips and worktrips for every trip	Data	Other	Fixed	1.3
36	REGION number discrepancy	Data	Other	Fixed	1.4
37	87 appropriate skips on FUTURESURVEY	Data	Other	Fixed	1.5
38	UPCODING of open end responses	Process	Other	Step was skipped for Pilot deliverable due to time constraints; will be done for main	1.6
39	EDUC has a lot of -9 (Not ascertained responses).	Instrument	Logic	Fix made to skip logic for main survey to ask this appropriately	1.7

Item #	Change Description	Survey Element	Change Type	Note	Appendix C Location
40	TCOFF - incorrect appropriate skip count?	Data	Other	We're asking more people this question only because we choose not to ask people who work at home this question if they are also self-employed.	1.8
41	CPLOG and HVLOG results illogical based on skip logic	Instrument	Logic	Fixed	1.9
42	LOCTYPE codes of 4 for all other locations besides Home, Work, or School?	Data	Other	Codes fixed to match codebook - we had used old codes from another survey	1.10
43	ADDSTRT blank but address appears in the city field	Data	Other	Fixed	1.11
44	Use of intersections for geocoding?	Data	Other	Geocodes with 0 distances from street network are reviewed and adjusted when possible	1.12
45	Definition of samplemp, samplema, and samplez9	Data	Other	Added codes to the data elements	1.13
46	HHVEH_USED has 513 not ascertained results	Instrument	Logic	Error in logic fixed for main survey	1.14
47	LNGTRIP implemented fully for spring?	Data	Other	This was fully implemented for spring data collection	1.15
48	Meaning of 'google_maps_missing_type' geocoding type?	Data	Other	Added a check to review these cases	2.1
49	TPURP and TPURP2 '97 - Other' records to be up-coded?	Data	Other	Step was skipped for Pilot deliverable due to time constraints; will be done for main survey	2.2
50	Place fields not updated with address data from location file	Data	Other	Deliverable generation had an update issue; fixed	2.3
51	GPS_ASSOC filled with '?'	Data	Other	A typo in the deliverable generation code caused this - fixed	2.4
52	Are time and distance checks being run?	Data	Other	These are being run	2.5
53	Discrepancy between GPS Place count and Place count	Data	Other	This is correct - see Appendix C, item 2.6, for details	2.6
54	Households with more places in Place File than in the GPS Place file?	Data	Other	This is correct - see Appendix C, item 2.6, for details	3.1

Item #	Change Description	Survey Element	Change Type	Note	Appendix C Location
55	Second case like 3.1	Data	Other	This is correct - see Appendix C, item 2.6, for details	3.2
56	Lat-Long disagreement between Place File and GPS Place File?	Data	Other	Error found and fixed in deliverable generation code	3.3
57	Presence of null lat long in the GPS Place File?	Data	Other	This is expected for home locations where no GPS existed	3.4
58	Trip Rates very close between GPS and non-GPS?	Data	Other	Westat believes this is primarily a factor of the small GPS sample size	3.5
59	Transit trips missing access or egress trips	Data	Other	Based on client input, transit checks were turned off; consensus was that incomplete transit tours are better than no transit tours	3.6
60	CHGADDBOX variable has '?'	Data	Other	Should have been coded as appropriate skip - ? Due to no PO Box households responding in the pilot	4.1
61	Cases with more than 6 decimal places in LONGITUDE and LATITUDE location file	Data	Other	DBF formatting issue	4.2
62	disallow sample_z9 geocode type	Data	Other	Enhancements to base geocodes, and getting updates to low precision geocodes, should resolve this for the main survey	4.3
63	No geocode type in the place file?	Data	Other	This exists in the location file, which can be joined to the place file	4.4
64	ACTDUR and TRAVTIME are real numbers; should be integer	Data	Other	DBF formatting issue will be resolved with casting data types	4.5
65	DISTANCE field has too many decimal places (should be 2)	Data	Other	DBF formatting issue will be resolved with casting data types	4.6
66	Longitude and Latitude field has too many decimal places (should be 6) in place file	Data	Other	DBF formatting issue will be resolved with casting data types	4.7
67	blank values for CITY, STATE, ZIP and COUNTRY in place file	Data	Other	Will be corrected for main data deliveries	4.8
68	Too many decimal places in ACTDUR, TRAVTIME, DISTANCE, LONGITUDE, and LATITUDE fields in GPS place file	Data	Other	DBF formatting issue will be resolved with casting data types	4.9

Item #	Change Description	Survey Element	Change Type	Note	Appendix C Location
69	MODE field filled with codes 0–5 (What do these mean?) in GPS Place file	Data	Other	Not able to see 0s in the deliverable file; only finding 1-97	4.10
70	One record with blank geocode type in Long Distance File	Data	Other	Will be corrected for main data deliveries	4.11
71	Too many decimal places in LATITUDE and LONGITUDE fields Long Distance File	Data	Other	DBF formatting issue will be resolved with casting data types	4.12
72	27 records with speeds greater than 95 mph	Data	Other	most of these are due to times being rounded on very short trips	5.1
73	7 trips with distance = 0	Data	Other	reference 5.3	5.2
74	101 records with speeds less than 3 mph and 265 with speeds less than 5 mph	Data	Other	We do review these and send to research as needed	5.3
75	Instances of 'I don't know' responses from questions like 'SEX', 'AGE', or 'MODE'	Data	Other	This is expected from web survey participants - high non-responses households will be eliminated	S1
76	Conflation of 0 versus Null	Data	Other		S2
77	Review of SCHOL 97, 98, 99	Data	Other		S3
78	DD Degree meaning?	Data	Other	Doctor of Divinity	S4
79	Using a whole number in place of N/A for appropriate skips?	Data	Other		S5
80	High number of appropriate skips. Can these be reduced?	Data	Other		S6
81	Code book display using % is confusing	Data	Other		S7
82	Request for regional summary to be added to codebook	Data	Other		S8
83	Refusal rate difference is high for Income (comparing MDOT to SEMCOG)	Data	Other		S9
84	Up-coding of trip purposes?	Data	Other	Step was skipped for Pilot deliverable due to time constraints; will be done for main survey	S10

Appendices

Appendix A. MDOT Weekly Report

Appendix B. SEMCOG Weekly Report

Appendix C. Log of Comments and Main Survey Updates

Appendix D. Public Awareness Materials

Appendix E. Recruitment Instruments

Appendix F. Retrieval Instruments

Appendix G. Survey Participant Materials

The following pages contain the complete list of participant materials, as follows:

Advance Materials and Invitations - MDOT

Advance Materials and Invitations - SEMCOG

Survey Packet Materials

GPS Materials

Spanish Language Materials

Arabic Language Materials

Travel Logs and Example Logs

Appendix H. Text and Email Reminders

Appendix YY

Spring Data Collection Memo

Westat®

An Employee-Owned Research Corporation®

Date: September 18, 2015

To: Don Mayle and Karen Faussett, Project Managers
Michigan Department of Transportation (MDOT)

Tom Bruff, Project Manager
Southeast Michigan Council of Governments (SEMCOG)

From: Jesse Casas, Martha Wilaby, and Jeremy Wilhelm, Westat

Subject: Spring Interim Report

---

1 Introduction

This memorandum provides a summary of the spring 2015 data collection effort for MTC III / STC 15 including an overview of the spring data collection goals and actual outcomes; an assessment of data quality, specifically item non-response levels, a geocoding assessment, and the quality of the Global Position System loggers and prompted recall (GPS+PR) survey data; a discussion about survey methods including analysis between web and computer assisted telephone interview (CATI) respondents and the GPS deployment process; and finally a summary of the fall data collection adjustments and recommendations based on spring sampling analysis, any adjustments to the survey instruments, project website, and public outreach efforts.

Following a successful pilot test, Westat fielded the MTC III / STC 15 spring data collection from April to June of 2015. A higher than anticipated response rate was both rewarding and a challenge.

2 Spring Data Collection Overview

2.1 General Staffing and Spring Schedule

Westat staffed for a spring data collection planned around an address-based sample of 190,000 households with three mail contact attempts: an invitation letter, a first reminder postcard after seven days, and a second reminder postcard after 14 days. Households self-recruited by web, with a toll free line available if completing the survey by telephone was preferred. Travel packets including a cover letter, travel logs, example logs, long distance travel logs, and a $1 primer, were prepared and sent to arrive a few days before the assigned travel date. Retrieval was also completed by web or telephone as requested. We started with a goal to complete 7,972 households.

Westat staffed a project-specific help desk to answer incoming calls, complete surveys by phone, make reminder calls, make equipment return reminder calls, and to conduct follow-up research calls for questions about reported travel. We used our Telephone Research Center (TRC) for outbound retrieval calls. Web non-responders were contacted by phone two days after their travel day.

Global Position System loggers and prompted recall surveys were utilized for 10% of households who were selected to be part of the GPS sample. GPS households received the standard travel packet materials along with GPS devices for each eligible person, GPS device use and return instructions, and a pre-paid return FedEx Pak. For GPS add-ons, we paid $25 per GPS-instrumented person (for all HH members between 16 and 75) in lieu of the $20 household incentive. GPS households did not receive the $1 primer.

The MI Travel Counts toll free phone number rang into the project help desk, which was staffed Monday through Friday from 9am to 9pm, Saturday from 10am to 6pm, and Sunday from 2pm to 9pm ET. Throughout the spring, a staff of 8 to 15 worked the help desk depending on workload. Staff were trained April 6-8, 2015 and the help desk opened on April 9. The planned end date for the help desk was June 13, but with higher response rates than expected we extended the help desk to June 28 to accommodate more respondents and research cases.

Westat's Telephone Research Center (TRC) made outbound calls for the retrieval surveys. TRC staffed 15-20 interviewers throughout the spring data collection effort. Hours of operation matched the help desk. While help desk staff were all located on the Rockville campus, TRC interviewers were remote workers.

All of our fulfillment and data processing staff were located in the Atlanta office. There were 16 people working on fulfillment and/or data processing tasks in the spring. Most staff had previous experience and they were cross-trained in at least two of the primary tasks. Some staff provided support for all of the tasks at varying times during data collection. The number of staff assigned to a given task was based on the volume of households recruiting into the survey and completing the retrieval survey.

The first travel date of the spring data collection was April 20, 2015. The invitation letters were mailed by US mail in four batches. The first 100,000 letters were mailed on April 6, 2015, followed seven days later by a reminder/thank you postcard to everyone. A second reminder postcard was sent 14 days after the initial invitation letter and seven days after the first reminder postcard. This second postcard was sent to only those addresses that had not yet responded to the survey. The second batch of 25,000 of letters was sent on April 14, 2015. The third batch of 25,000 of letters was sent on April 21, 2015. The fourth batch of 40,000 letters was sent on April 28, 2015. Participation incentives for non-GPS households ($20) were paid by check through US mail.

Households that chose to have the travel logs mailed to them were assigned a travel date at least 10 days in the future so the travel logs could be prepared and mailed to arrive a few days before the travel date. Those who opted to print their own materials were assigned the next available date. The last travel date was June 4, 2015.

2.2 Original Recruitment Sampling Goals Versus Actual

Table 1 shows a comparison of targeted recruitment numbers and actual results for spring data collection by sponsor. Westat began data collection with spring recruitment goals of 12,859 overall households (8,643 in MDOT and 5,718 in SEMCOG, with an overlap of 1,502 MDOT-funded households in the SEMCOG counties). In actual numbers, a total of 16,974 households agreed to participate in the survey with 12,407 in MDOT, 6,540 in SEMCOG, and 1,973 in the overlap.

Table 1. Recruitment Targets and Actual Results

Sponsor	Target	Actual	Percent of Target
MDOT Funded in SEMCOG Area	1,502	1,973	131.40%
SEMCOG Funded in SEMCOG Area	4,216	4,567	108.30%
MDOT Funded in MDOT Area	7,141	10,434	146.11%
Total	12,859	16,974	132.00%
All MDOT Households	8,643	12,407	143.55%
All SEMCOG Households	5,718	6,540	114.40%

2.3 Original Retrieval Sampling Goals Versus Actual

Table 2 shows results for targeted and actual retrieval numbers for spring data collection by sponsor. The retrieval goals at the beginning of spring data collection were 5,358 MDOT-funded households (931 in the SEMCOG region) and 2,614 SEMCOG-funded households for a total of 7,972 completes. The actual retrieved count of households were 8,113 in the MDOT-funded sample (1,298 in the SEMCOG region) and 2,947 SEMCOG-funded households for a total of 11,060.

Table 2. Retrieval Targets and Actual Results

Sponsor	Target	Actual	Percent of Target
MDOT Funded in SEMCOG Area	931	1,298	139.40%
SEMCOG Funded in SEMCOG Area	2,614	2,947	112.70%
MDOT Funded in MDOT Area	4,427	6,815	153.94%
Total	7,972	11,060	138.70%
All MDOT Households	5,358	8,113	151.40%
All SEMCOG Households	3,545	4,245	119.70%

2.4 Expected Individual and Composite Response Rates Versus Actual

Based on the pilot survey, an overall participation rate of 4.875% was anticipated for the spring data collection effort (7.5% recruitment rate * 65% retrieval rate). Table 3 shows the performance of the fall sample for each modeling area/county with each area's performance measured against the anticipated 4.875% participation rate. Areas or counties that underperformed are shaded orange.

Table 3. Recruitment, Retrieval, Participation Rates versus Expected Spring Rates by Model Area

Sample Area	Rec. %	Ret. %	Part. %	+/- 4.875%
Statewide Model				
Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County (WATS)	7.2%	62.5%	4.5%	-0.4%
Southern Michigan Rural	9.5%	66.4%	6.3%	1.4%
Northern Michigan Rural	10.8%	66.5%	7.2%	2.3%
Small Cities	9.9%	64.2%	6.3%	1.5%
Statewide Model Total	8.9%	64.8%	5.8%	0.9%
Urban Model Areas				
Grand Valley Metropolitan Council (GVMC)	11.9%	66.1%	7.8%	3.0%
Tri-County Regional Planning Commission (TCRPC)	11.2%	71.0%	8.0%	3.1%
Genesee County Metropolitan Planning Commission (GCMPC)	8.3%	61.0%	5.0%	0.2%
Great Lakes Bay Region (GLBR)	10.3%	65.9%	6.8%	1.9%
Washtenaw Area Transportation Study (WATS)	9.7%	71.8%	7.0%	2.1%
Kalamazoo Area Transportation Study (KATS)	10.8%	62.9%	6.8%	1.9%
West Michigan Metropolitan Transportation Planning Program (WestPlan)	9.3%	63.3%	5.9%	1.0%
Jackson Area Comprehensive Transportation Study (JACTS)	8.8%	62.1%	5.5%	0.6%
Twin Cities Area Transportation Study (TwinCATS) and Niles/Buchanan/Cass Area Transportation Study (NATS)	9.3%	63.1%	5.9%	1.0%
Macatawa Area Coordinating Council (MACC)	10.6%	65.7%	7.0%	2.1%
Battle Creek Area Transportation Study (BCATS)	8.3%	64.6%	5.4%	0.5%
Traverse City (TVC)	12.7%	66.8%	8.5%	3.6%
Urban Model Areas Total	10.2%	65.7%	6.7%	1.8%
SEMCOG Counties				
East Wayne	6.4%	54.5%	3.5%	-1.4%
West Wayne	7.2%	65.1%	4.7%	-0.2%
Oakland	7.7%	66.4%	5.1%	0.2%
Macomb	6.4%	65.1%	4.2%	-0.7%
Washtenaw	9.7%	71.8%	7.0%	2.1%
Monroe	7.1%	65.1%	4.7%	-0.2%
St. Clair	8.0%	66.4%	5.3%	0.4%
Livingston	9.9%	66.4%	6.5%	1.7%
SEMCOG Counties Total	7.5%	64.9%	4.9%	0.0%
Grand Total	8.9%	65.2%	5.8%	0.9%

2.5 GPS and Prompted Recall Recruitment and Retrieval Expectations and Actual Performance

The purpose of the subsample of GPS+PR households is to estimate trip rate correction factors to adjust for trip underreporting in the non-GPS survey sample. GPS+PR provides a way for participants to confirm GPS trips and details, or to add trips not captured by GPS, while viewing their GPS data and their travel logs together. This approach allows for a travel day collection that leverages the strengths of the GPS technology and the strengths of the travel log methodology while mitigating the weaknesses of each.

The overall goal for the survey is to obtain a 10% subsample of complete GPS+PR households. The target for spring data collection was to complete 38% of the overall goal. There were no set targets at the sampling area or sponsor level. For informative purposes the actual recruit values for sponsor are included in Table 4.

Table 4. Recruitment Results for GPS+PR Households

Sponsor	Recruited
MDOT	923
SEMCOG	543
Subtotal	1,466
Overlap	-168
Final Total	1,298

While the actual total recruit volume of non-GPS households exceeded the target for the spring, the GPS recruitment remained consistent throughout the data collection by setting the GPS household daily limit at 46. See section 4.2 for GPS deployment methods.

Table 5 shows a comparison of the expected retrieval targets and the actual retrieval rates experienced in the spring. While the recruitment rate did not reflect proportionally to the non-GPS recruitment rate, the GPS household retrieval rate was reflective of the higher than expected retrieval rate in the overall spring data collection effort. Of the 1,298 households recruited to participate in the GPS subsample, 953 completed the retrieval survey, resulting in a 73% retrieval rate and 747 were subsequently cleared to be delivered for a 58% delivered rate. More discussion about the differences between retrieval and delivered rates is presented in Section 3.3.

Table 5. Retrieved and Delivered Results for GPS+PR Households

Sponsor	Target	Retrieved	Delivered	Percent of Target
MDOT	--	683	523	--
SEMCOG	--	389	313	--
Subtotal	--	1,072	836	--
Overlap	--	-119	-89	--
Final Total	800	953	747	93%

3 Data Quality Assessment

3.1 Item Non-Response

The Pilot Survey Report assessed item non-response as a means for determining the effectiveness of the survey questions. Table 6 contains a list of questions which received high item non-response in the pilot, and compares the pilot results with spring results. It also includes additional records where the non-response rate was at or above 1%. Westat does not recommend any changes to address the item non-response that was observed in the spring.

Table 6. Item Non-Response

Variable	MDOT Pilot	MDOT Spring	SEMCOG Pilot	SEMCOG Spring
Age (and age range)	0.2%	0.2%	1.2%	0.3%
Gender	0.6%	0.8%	2.0%	1.0%
License	0.5%	0.4%	1.5%	0.4%
Employed	0.1%	0.6%	1.5%	0.8%
Employer description	2.8%	1.6%	6.5%	1.5%
Industry	0.8%	1.4%	2.5%	1.5%
Telecommuting offered	7.4%	5.7%	7.9%	5.2%
Level of school	1.0%	0.3%	5.7%	0.4%
Home ownership status	4.5%	0.6%	5.7%	3.0%
Future survey participation	11.2%	10.3%	6.0%	9.0%
Trip purpose	0.5%	0.3%	1.4%	0.2%
Reason for longer than usual trip	-	4.9%	-	4.6%
Educational attainment	-	1.2%	-	1.6%
Number of jobs	-	1.0%	-	1.2%
Trip length longer than usual	-	0.9%	-	0.9%
Complete log	-	0.9%	-	1.2%
Student status	-	0.8%	-	0.9%
Work location	-	0.8%	-	0.9%
Reason for no license	-	0.7%	-	0.9%

3.2 Geocoding Assessment

Westat assessed the quality of geocoded locations in the spring data compared to the requirements found in the RFP. Table 7 provides each requirement from the RFP alongside the assessment of the data that has been collected to date. It is our conclusion based on the following assessment that the requirements set forth in the RFP have been met or exceeded, and that the geocoding in the spring effort was of high quality.

Table 7. Geocoding Assessment

RFP Requirement	Spring Assessment
1) All geocoded points will be provided in longitude and latitude.	- Passes requirement - All locations include an XY coordinate.
2) For points geocoded to longitude and latitude, the hierarchy of preferred spatial scales is: a) physical street address, then b) nearest intersection	- Passes requirement - All delivered locations have a minimum precision of 'nearest intersection'.
3) The following targets are to be met: a) 99% or more of home addresses will be geocoded to longitude and latitude. b) 95% or more of all school and work locations will be geocoded to longitude and latitude. c) 90% or more of other stops/locations will be geocoded to longitude and latitude.	- Passes requirements - 100% of 59,012 addresses have been geocoded to longitude and latitude
4) Offsets are to be a minimum[1] of 25 feet.	- 5,609 out of 59,012 locations (9.5%) failed this criteria and were closer than 25' to the MDOT road network centerline shapefile. - 2,482 are geocoded to 'nearest intersection' - The remaining 3,127 locations account for 5.3% of all locations o 2.1% are within 10' o 3.5% are within 15' - All have been reviewed by data analysts and re-geocoded using Bing, Mapquest, or Open Street Map where possible - 1,797 of these locations are in rural areas where spatial coverage and geocoding precision is presumably lower

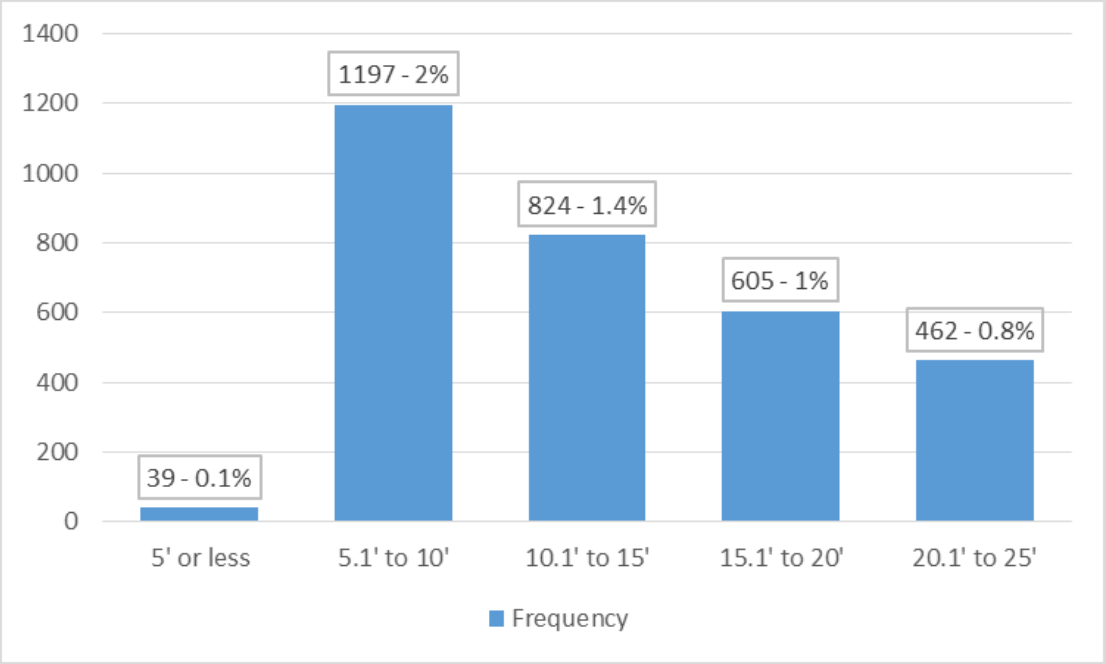
[1] The RFP actually reads 'maximum of' but it has been clarified in prior meetings that it should actually be a minimum of 25'. The rationale behind this requirement is that locations closer than 25' from the road network (which is typically the street centerline) may be too close to the road and perhaps too close to a TAZ boundary to allow for a confident allocation of the location to the correct TAZ.

Table 7. Geocoding Assessment (Continued)

RFP Requirement	Spring Assessment
5) For locations that are not automatically geocoded, the consultant will develop a process for online and map checks to manually geocode those locations.	All locations were geocoded in real-time during recruitment and retrieval surveys. Additional reviews were conducted to manually adjust locations that violated basic speed checks, network proximity checks, had a missing street number, or had a level of geocoding precision lower than 'nearest intersection'. These were reviewed by data analysts and re-geocoded using Bing, Mapquest, or Open Street Map resources when possible. If unresolved, cases were sent to research for follow-up calls. Cases that could not be clarified were failed and not delivered.
6) Only after the manual geocoding options are exhausted will the location be deemed non-geocodable.	See item 5 assessment.
7) A household will be considered incomplete if 25% or more of its locations are non-geocodable.	No households were delivered which had 25% or more of non-geocodable locations. In fact, all households delivered had 100% of locations geocoded.

Figure 1 shows the distribution of locations that were within 25 feet of the street centerline. It excludes locations geocoded to 'nearest intersection' and includes locations regardless of rural/urban classification. The locations in the analysis are for all sample regardless of the sponsor.

Figure 1. Distribution of Locations Offset within 25' of the MDOT Road Network Shapefile[2]



3.3 GPS and Prompted Recall Households

Table 8 shows a summary of GPS households recruited, returned, released, completed, and scheduled for delivery by sampling area during the spring data collection. GPS sample retrieval rates were much higher than anticipated overall and in most of the sampling areas. East Wayne had the lowest retrieval rate at 53.1 percent. The overall retrieval rate was 73 percent with the highest (81 percent) occurring in both Southern Rural Michigan and Grand Valley.

During the data quality control process, the 953 GPS households that completed retrieval were reviewed. Of these, 20 percent were disqualified from delivery for various reasons, including cases where a participant changed the data in a fashion that violated basic speed checks by moving a GPS-based location away from the GPS derived geocode, or cases where a participant ignored the times associated with a GPS-derived location and entered new arrival or departure times that violated basic speed checks. This resulted in a delivered rate of 58 percent, which is 2 percent lower than the anticipated rate of 60 percent. The shortfall from the spring will be recovered in the fall.

[2] This figure excludes cases matched to the nearest intersection.

Strategies for improving data quality from participants have been added for the fall data collection including prompts that warn participants when they are about to change the GPS data in a way that renders the results illogical and unusable.

Note: There were no GPS+PR goals by sampling area.

Table 8. GPS Recruitment, Deployment, and Retrieval Results – MDOT Modeling Areas

Sample Area	Recruited	Returned	Released	Completed	Retrieval Rate	Delivered	Delivered Rate
Statewide Model							
Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County (WATS)	102	80	80	67	66%	52	51%
Southern Michigan Rural	63	59	59	51	81%	40	63%
Northern Michigan Rural	81	70	69	56	69%	41	51%
Small Cities	57	46	46	38	67%	27	47%
Statewide Model Total	303	255	254	212	70%	160	53%
Urban Model Areas							
Grand Valley Metropolitan Council (GVMC)	85	77	77	69	81%	53	62%
Tri-County Regional Planning Commission (TCRPC)	56	53	53	45	80%	35	63%
Genesee County Metropolitan Planning Commission (GCMPC)	52	39	39	34	65%	28	54%
Great Lakes Bay Region (GLBR)	60	54	54	46	77%	32	53%
Washtenaw Area Transportation Study (WATS)	66	61	60	52	79%	37	56%
Kalamazoo Area Transportation Study (KATS)	48	43	41	36	75%	28	58%
West Michigan Metropolitan Transportation Planning Program (WestPlan)	42	37	37	32	76%	24	57%
Jackson Area Comprehensive Transportation Study (JACTS)	46	39	38	31	67%	29	63%
Twin Cities Area Transportation Study (TwinCATS) and Niles/Buchanan/Cass Area Transportation Study (NATS)	41	37	36	33	80%	25	61%
Macatawa Area Coordinating Council (MACC)	29	25	25	21	72%	15	52%
Battle Creek Area Transportation Study (BCATS)	38	32	32	27	71%	21	55%
Traverse City (TVC)	57	50	50	45	79%	36	63%
Urban Model Areas Total	620	547	542	471	76%	363	59%
Combined Total	923	802	796	683	74%	523	57%

Table 8. GPS Recruitment, Deployment, and Retrieval Results – MDOT Modeling Areas (Continued)

Sample Area	Recruited	Returned	Released	Completed	Retrieval Rate	Delivered	Delivered Rate
SEMCOG Counties							
East Wayne	64	45	43	34	53%	22	34%
West Wayne	98	83	83	70	71%	54	55%
Oakland	131	113	113	95	73%	81	62%
Macomb	68	59	58	50	74%	46	68%
Washtenaw	66	61	60	52	79%	37	56%
Monroe	30	24	24	23	77%	19	63%
St. Clair	31	28	28	23	74%	20	65%
Livingston	55	48	47	42	76%	34	62%
SEMCOG Total	543	461	456	389	72%	313	51%
MDOT Total	923	802	796	683	74%	523	57%
Minus Overlap	-168	-141	-140	-119	--	-89	--
Grand Total	1,298	1,122	1,112	953	73%	747	58%

Table 9 shows a comparison of the trip rates from log-only households and GPS households. The pilot results raised concerns that the difference in trip rates were negligible. For the spring data collection effort, we can see that the rates are noticeably higher in the GPS sub-sample. This is more aligned with our experience on prior GPS surveys.

Table 9. Comparison of Log and GPS Person Trip Rates

Sample Area/County	Log-only Households	Log-only Persons	Log-only Trip Rates	GPS Households	GPS Persons	GPS Trip Rates
Statewide Model						
Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County (WATS)	3,377	7,600	3.78	276	649	4.58
Southern Michigan Rural	617	1,466	3.48	40	107	5.38
Northern Michigan Rural	645	1,395	3.64	41	97	6.01
Small cities	471	993	4.15	27	70	6.36
Urban Model Areas						
Grand Valley Metropolitan Council (GVMC)	644	1,497	3.82	53	137	4.58
Tri-County Regional Planning Commission (TCRPC)	538	1,137	3.86	35	83	4.87
Genesee County Metropolitan Planning Commission (GCMPC)	316	669	4.00	28	72	4.18
Great Lakes Bay Region (GLBR)	449	964	4.05	32	76	5.33
Washtenaw Area Transportation Study (WATS)	441	974	4.19	37	109	4.14
Kalamazoo Area Transportation Study (KATS)	436	1,022	4.01	28	63	4.06
West Michigan Metropolitan Transportation Planning Program (WestPlan)	303	702	3.76	24	52	5.33
Jackson Area Comprehensive Transportation Study (JACTS)	277	619	3.79	29	71	4.56
Twin Cities Area Transportation Study (TwinCATS) and Niles/Buchanan/Cass Area Transportation Study (NATS)	379	820	3.89	25	64	5.44
Macatawa Area Coordinating Council (MACC)	364	869	4.05	15	37	5.46
Battle Creek Area Transportation Study (BCATS)	291	574	3.99	21	46	5.57
Traverse City (TVC)	450	965	3.95	36	82	4.45
SEMCOG Sample						
East Wayne	430	859	3.90	22	49	4.24
West Wayne	782	1,702	3.87	54	132	4.50
Oakland	876	2,046	3.82	81	196	4.67
Macomb	491	1,086	3.60	46	108	4.53
Washtenaw	441	974	4.19	37	109	4.14
Monroe	224	547	3.74	19	36	5.83
St. Clair	264	606	3.68	20	42	4.52
Livingston	310	754	3.74	34	86	4.33
Grand Total	9,998	22,266	3.85	747	1,815	4.85

4 Methods Highlights

4.1 Web versus Computer Assisted Telephone Interview Respondent Demographic Analysis

The methodology used in MTC II differed from the current methodology of MTC III in key ways. This section examines the effect of one of them: the differences observed between web and Computer Assisted Telephone Interview (CATI) survey modes. In MTC II, the primary mode for survey completion was phone. Respondents could also choose to report their travel by web or mail their diaries. For MTC III, the primary mode for both recruit and retrieval is web. Respondents may also participate by phone (CATI). This section examines the difference between MTC III web and CATI survey modes.

In particular, the analysis is focused on a comparison of households which reported travel details by CATI or by Web, as well as averages regardless of mode used, versus 2012 ACS 5-year estimates for household income, household workers, household vehicles, and respondent age. In addition to socio-demographic comparisons, this section examines trip-making characteristics including household trip rates, trip purposes, and trip mode shares by reporting mode. Trip-making characteristics do not have ACS benchmarks for comparison.

A key point about this analysis is that CATI, representing the smaller share of the overall responses (roughly 21% of retrievals), is viewed as a modulator of the web responses, i.e., absent CATI responses, we see how Web responses alone would perform against ACS benchmarks, and by viewing all responses, we see the effect that CATI responses have on moving closer to, or further from, ACS benchmarks.

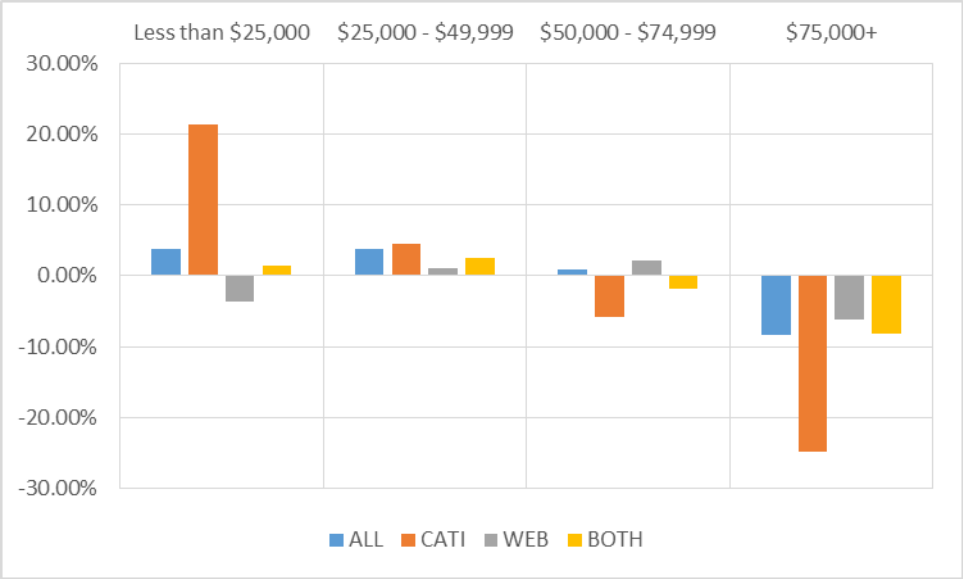
4.1.1 Socio-demographic Comparisons

The figures presented in this section include results for CATI-only responses, Web-only responses, all responses, and ACS. For the sake of clarity, households that utilized both CATI and Web modes evenly have been excluded. This is primarily due to the fact that such households match closely to the distributions for all modes combined, and therefore do not add new information to the results, and secondarily, because the readability of the figures was diminished with the inclusion of the additional data points.

4.1.1.1 Household Income

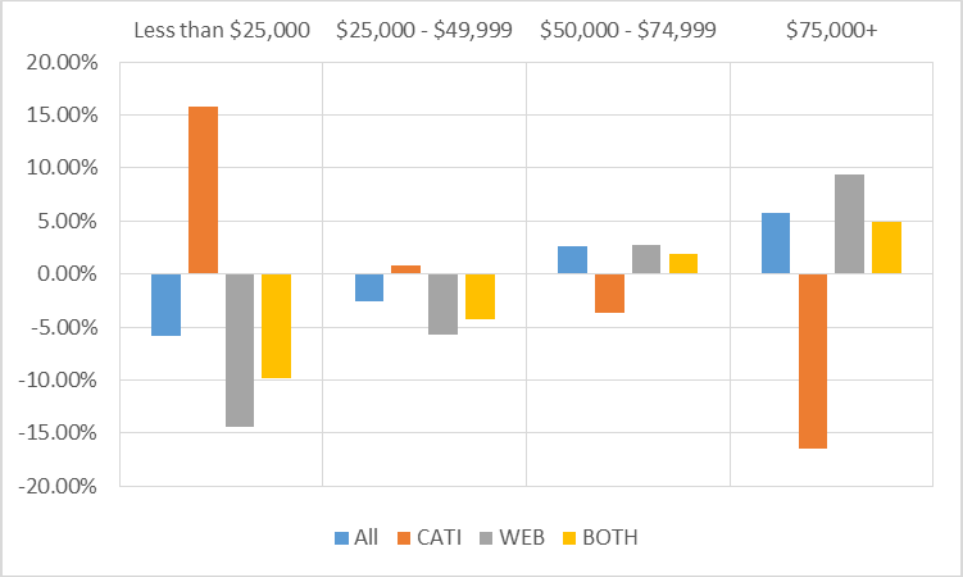
Household income was evaluated with quartile categories and by aggregating responses to originally reported values collected in the recruitment survey. Figure 2 shows clearly that the isolated web responses were closest overall to ACS benchmarks and that CATI responses in isolation were the furthest.

Figure 2. MDOT Sample – Percent Deviation from 5 Year ACS, 2012 – Household Income Quartiles



The same general pattern holds when analyzing SEMCOG samples, but it is notable that the effect in SEMCOG cases shows that the overall response is generally closer to ACS than Web in every quartile, suggesting that CATI responses for the SEMCOG sample generally improved the mix of responses (Figure 3).

Figure 3. SEMCOG Sample – Percent Deviation from 5 Year ACS, 2012 – Household Income Quartiles



4.1.1.2 Household Workers

Figure 4 shows the percent deviation from ACS for household workers reported in the MDOT sample. Here again, the pattern appears to show that presence of CATI responses, while well out of alignment with ACS, brings the total distribution of responses into closer alignment with ACS

benchmarks. Figure 5 shows that the same relationship holds for samples in the SEMCOG region.

Figure 4. MDOT – Percent Deviation from 5 Year ACS, 2012 - Household Workers

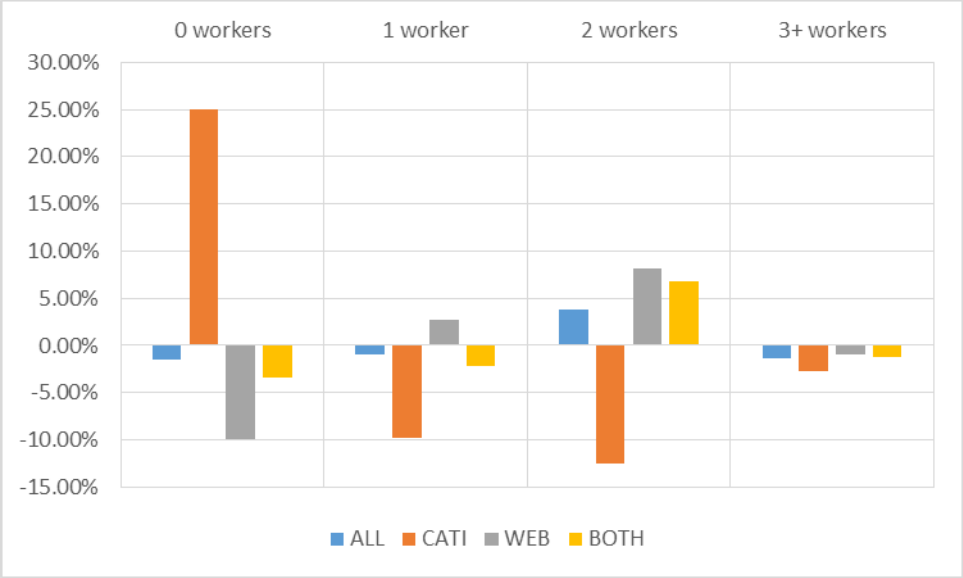
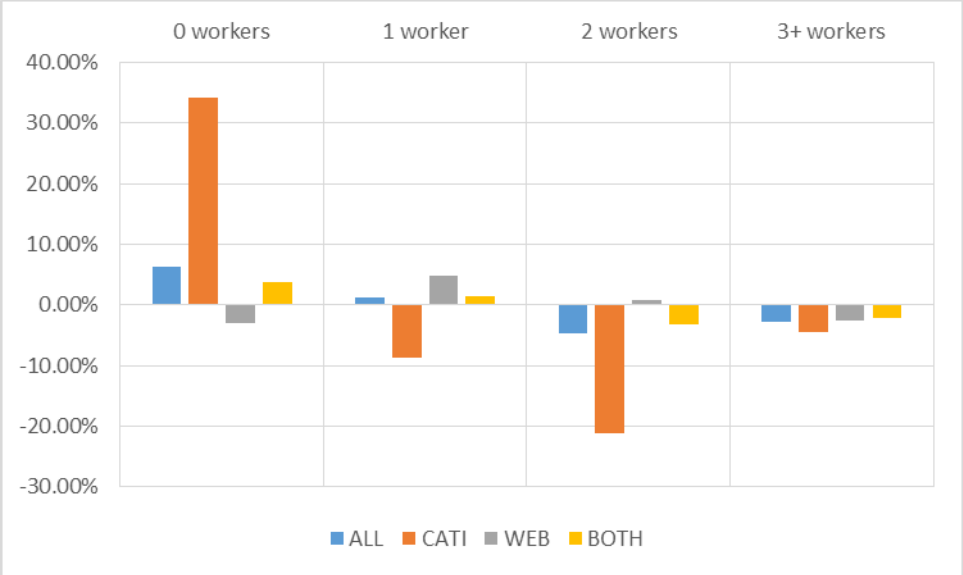


Figure 5. SEMCOG – Percent Deviation from 5 Year ACS, 2012 - Household Workers



4.1.1.3 Household Vehicles

The percent of household vehicles by mode are aligned closely to ACS benchmarks with CATI slightly over-representing 1 vehicles households, and Web surveys slightly over-representing 2-vehicle households, as seen in Figure 6 and Figure 7.

Figure 6. MDOT – Percent Deviation from 5 Year ACS, 2012 - Household Vehicles

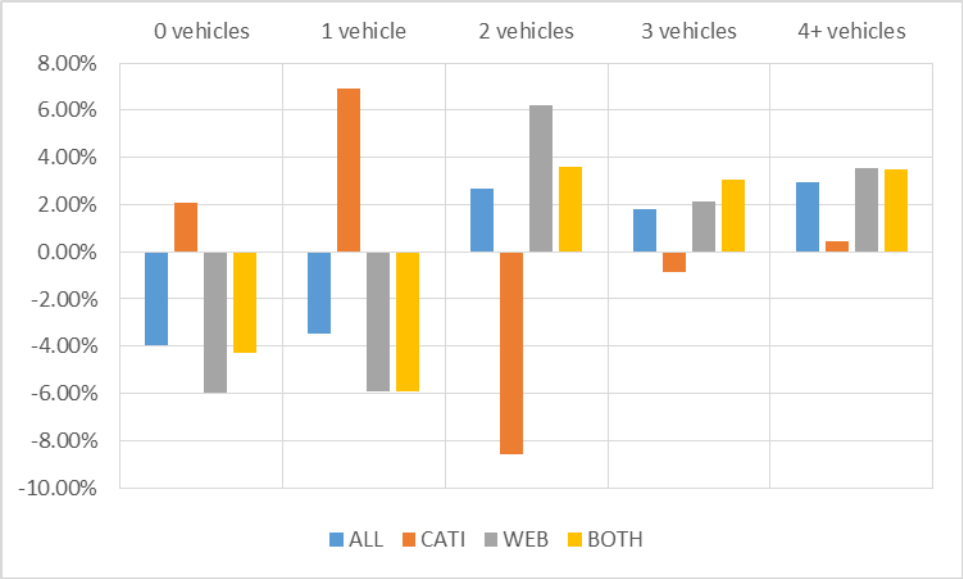
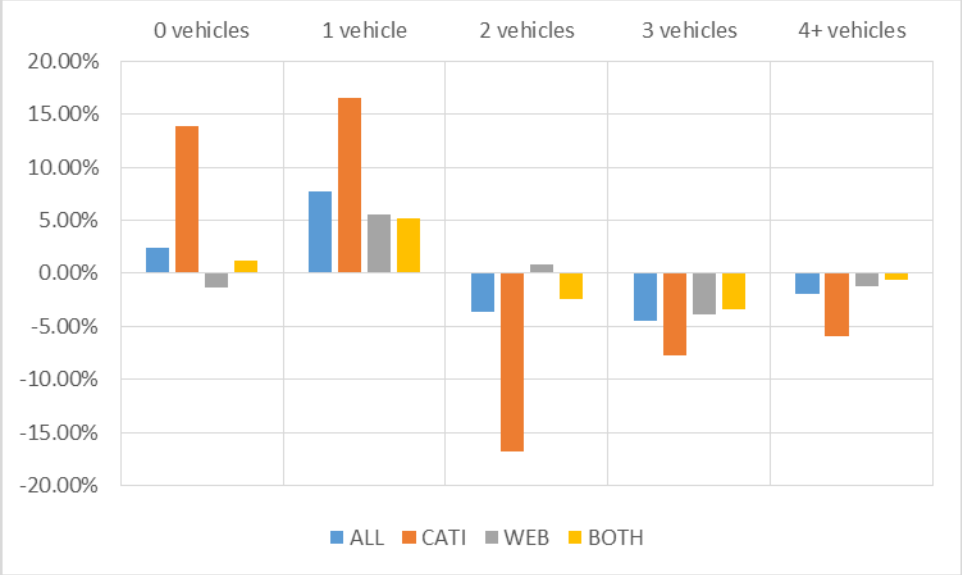


Figure 7. SEMCOG - Percent Deviation from 5 Year ACS, 2012 - Household Vehicles



4.1.1.4 Age Distributions

In Figure 8 and Figure 9, we see that web surveyed households are closer to ACS benchmarks for younger household members, and that CATI has a greater over-representation of older household members, especially in the 65-70 age range.

Figure 8. MDOT – Percent Deviation from 5 Year ACS, 2012 – Age Distribution

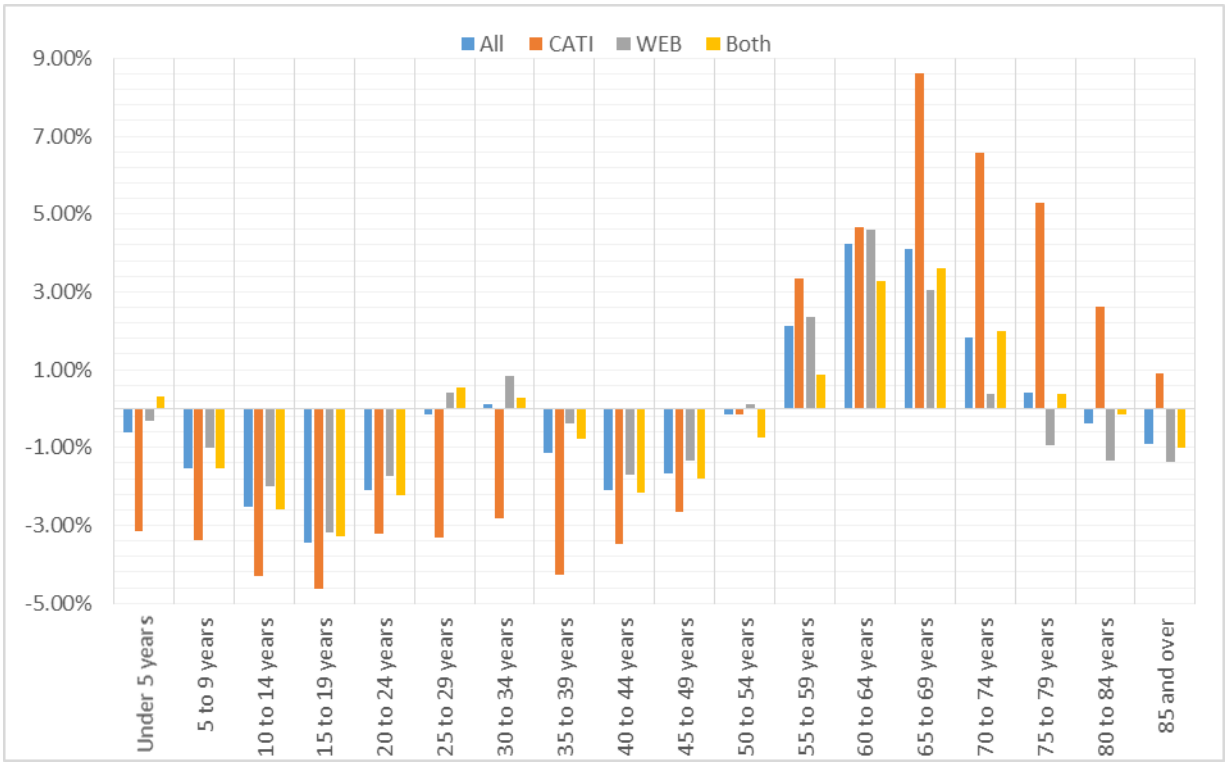
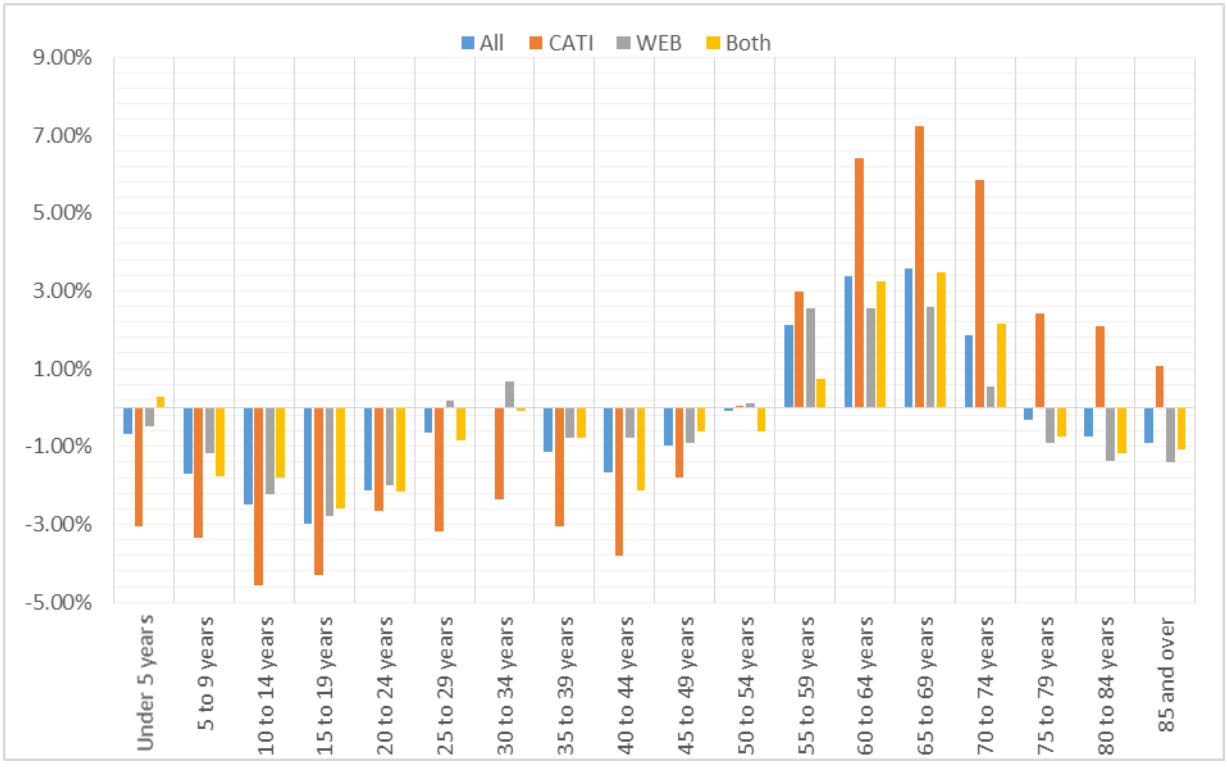


Figure 9. SEMCOG – Percent Deviation from 5 Year ACS, 2012 – Age Distribution



4.1.2 Travel Behavior Comparisons

The data shown in the comparison of trip rates has no ACS benchmark for comparison. However, comparing trip rates, trip purpose distributions, and mode share is illuminating for discerning differences in the reporting of households based on survey mode. Unlike in the previous sections, the figures in this section include households that reported using both modes.

4.1.2.1 Household Trip Rates

Looking at trip rates, Figure 10 and Figure 11 show that CATI-surveyed households have a larger share of households making 5 or fewer trips whereas these same households have a lower share making 8 or more trips. This is indicative of either less-active households, or households with fewer people, and perhaps both. Notably, households that used both survey modes equally have the highest proportion of participants with 18 or more trips. This is intuitive, as one might theorize that a household with a lot of travel would grow weary of online reporting and eventually call the helpdesk or receive a call from the helpdesk to assist in completing the survey.

Figure 10. MDOT – Household Trip Rates

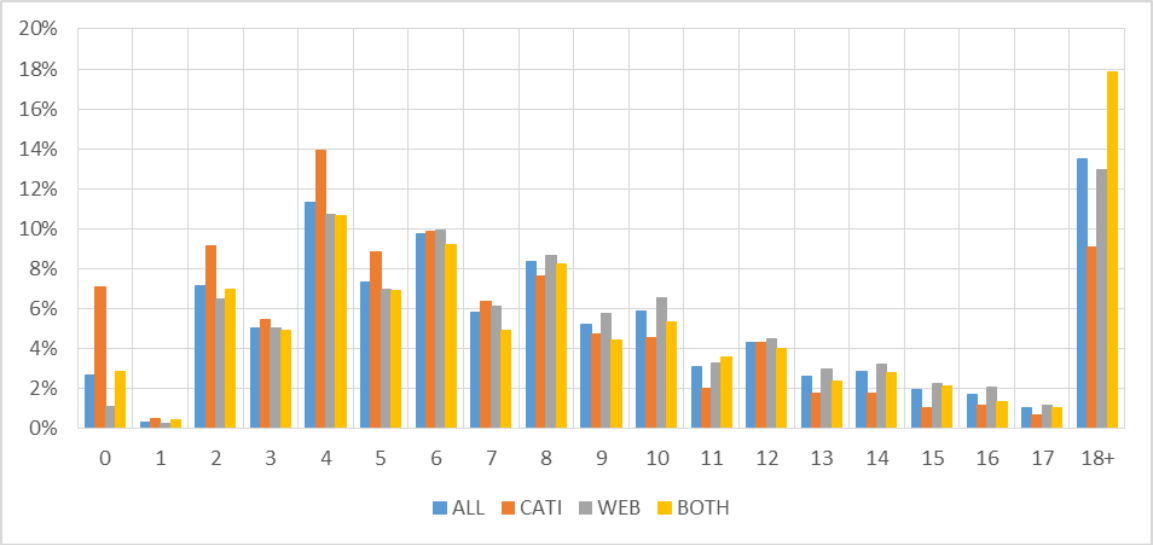
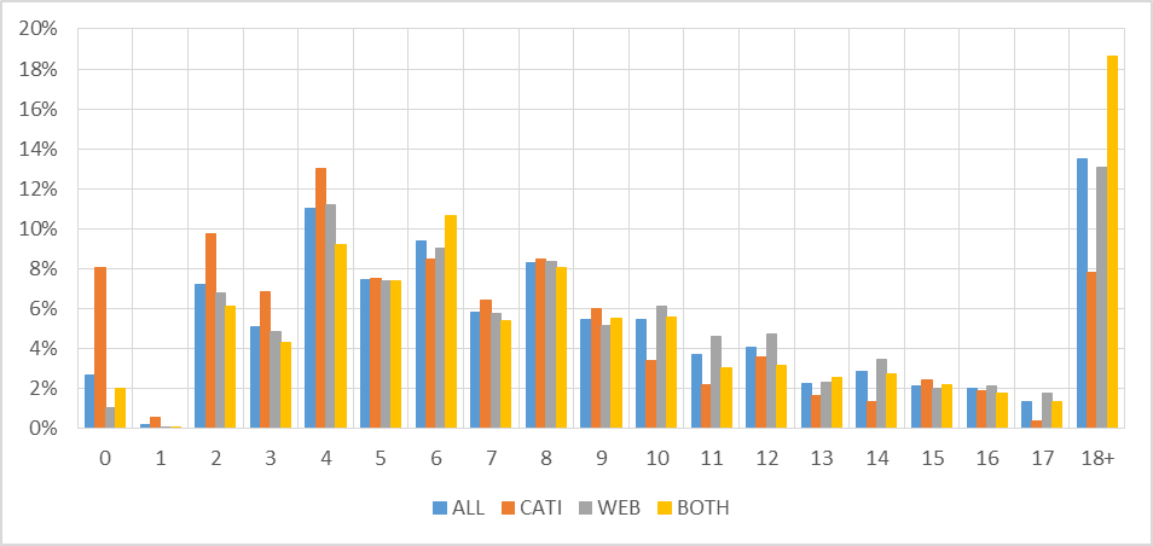


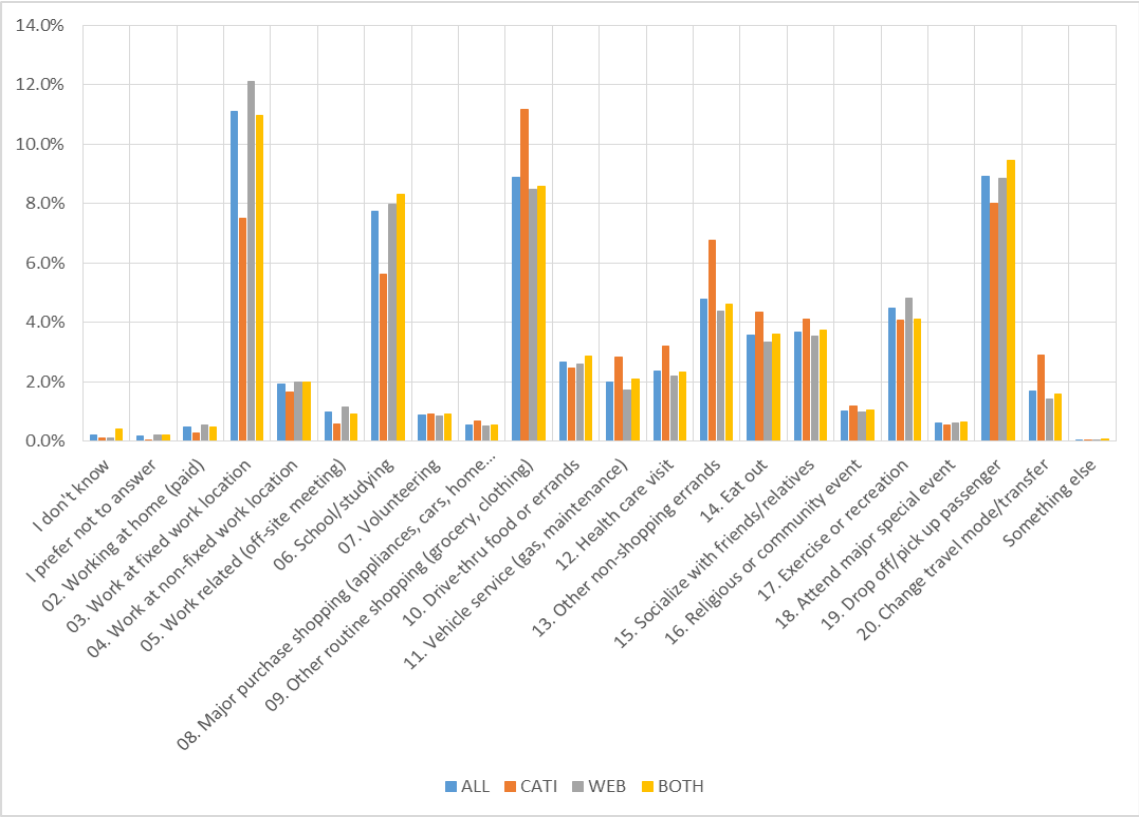
Figure 11. SEMCOG – Household Trip Rates



4.1.2.2 Trip Purpose

The trip purpose distributions in Figure 12 show that CATI-surveyed households tend to have a lower share of work and school purposes and a higher share of discretionary purposes, including making major purchases, servicing vehicles, and attending major events.

Figure 12. All Sample – Distribution of Trip Purposes



4.1.2.3 Mode Share

Figure 13 and Figure 14 show mode share by survey modes with the first showing auto modes isolated from the remainder (because including them with the other modes makes distributions illegible for the other modes). Auto modes are relatively equally represented regardless of survey mode. Looking at all other travel modes, CATI surveyed households show a higher share of non-motorized and transit-based trips.

Figure 13. All Sample – Auto Trips by Survey Mode

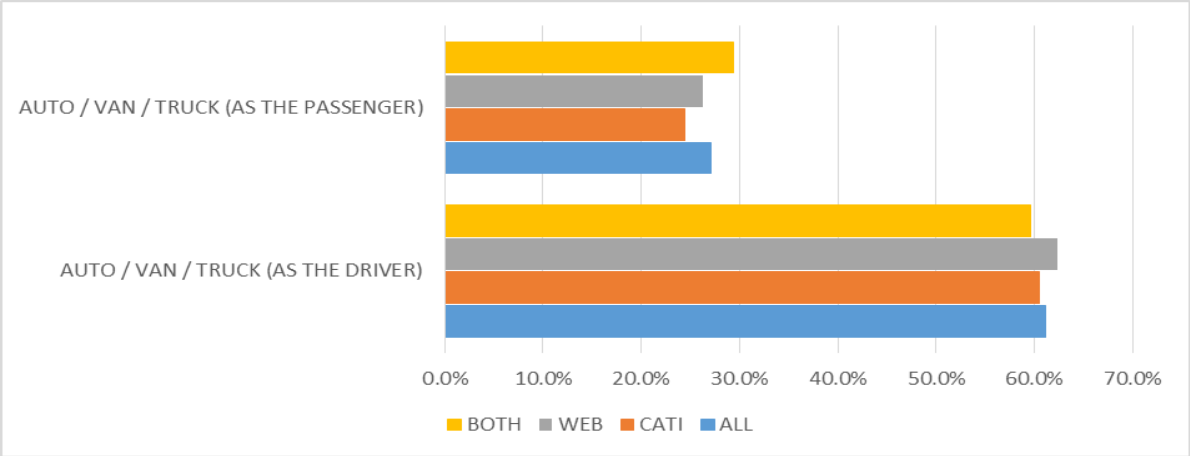
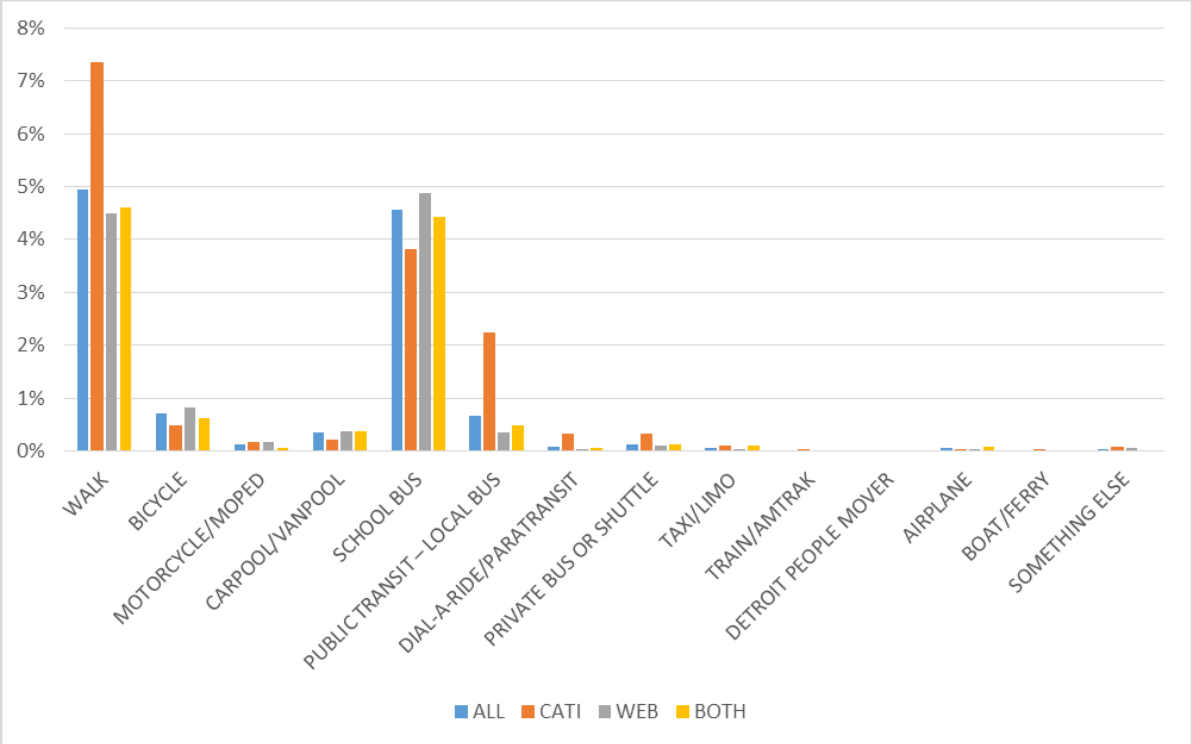


Figure 14. All Sample – Mode Share (Excluding Auto Modes)



4.1.3 Survey Mode Comparison - Concluding Observations

In general, the analysis suggests that CATI-surveyed households are either older in age makeup, smaller in size, or over-represented on the lower ends of the income spectrum. This suggests that CATI-surveyed households are critical for gaining proportional representations of some hard-to-reach household types. However, it is also apparent that CATI-surveyed households under-perform with multi-worker, multi-vehicle, and younger households; these are categories where web-surveyed households achieve distributions closer to ACS. While web-surveyed households tend to match fairly well to ACS benchmarks, this analysis supports the presumption that a balanced and effective household travel survey must include a thorough CATI effort in order to reach households of diverse socio-demographic and behavioral types.

4.2 GPS Deployment and Returns Assessment

By managing recruitment, Westat ensured that all GPS households were deployed on time. GPS deployment went smoothly throughout the spring data collection with only a couple of minor anticipated address and delivery exceptions.

After completing the recruitment survey, GPS devices were shipped to arrive a couple of days before the travel date. The GPS devices were used for three consecutive days with the starting day falling at some point on a Monday through Thursday. Devices were returned on the fourth (or in the case of Thursday households, the fifth) day via FedEx.

GPS households received a schedule of email and/or text equipment use and return reminders. They received a reminder every day of the three-day data collection to continue using the device. GPS households received reminders to return the GPS equipment on the 4th, 5th, and 7th days, then continuing every 3 days for 28 days following the 3rd day of data collection. Beginning at 30 days beyond the first travel date there were weekly reminders until the equipment was returned. They also received equipment reminder phone calls periodically. Households that did not return the GPS equipment within two weeks of their 3rd travel date received GPS equipment return request letters via mail.

Returned GPS equipment was processed expeditiously with the goal of GPS files being uploaded to a secure Westat server within 24 hours of receipt. This was to ensure that the remaining steps for retrieval completion could progress. After being uploaded, the GPS data was processed, and released to participants to review, confirm, adjust if necessary, and provide additional place details. A summary of GPS household deployment status is shown in Table 10.

Table 10. GPS Deployment Summary

Total Deployed	Total Returned	Still Deployed / Not Returned	% Not Returned
1,295	1,198	56	4.3%

The 4.3% of GPS household still deployed is acceptable and in fact is lower than Westat has experienced in previous GPS+PR studies. Westat will continue to use the reminder system to encourage GPS households to send the GPS equipment back quickly. By the end of the fall data collection, it is anticipated that the outstanding GPS equipment rate will be lower than the 4.3% experienced in the spring.

5 Fall Survey Sampling Adjustment Plan

5.1 Summary (interpretation) of Sampling Shortfalls

Westat used the response rates in the spring to predict the number of retrieval completes for various fall data collection scenarios. One scenario, referred to as Option 1, was to mail a sufficient number of letters in each model area so that the expected number of spring-plus-fall retrieval completes was equal to the sampling plan's target for the model area. The size of the fall mailing for Option 1 is approximately 190,000 households. This plan would satisfy the requirements of the scope of work.

However, because the spring response rates for some types of household were lower than those for other types of households, an expected outcome for Option 1 is that the number of completes for some types of households will be less than the sampling plan's target for that type of household. SEMCOG staff expressed concerns about having insufficient observations of some household types to allow for proper model estimation.

5.2 Plan to Reach Sampling Goals/Minimize Cell Target Shortfalls

To reduce anticipated shortfalls, Option 2 (Final) increases the number of mailed letters in selected model areas and, where it appears to further decrease shortfalls, increases the oversampling of the area's high-density stratum. The size of the fall mailing for Option 2 (Final) is approximately 235,000 letters, an increase of 45,000 mailed letters, which is expected to result in 1,900 to 2,000 additional retrieved households. It is also expected that implementing Option 2 (Final) instead of Option 1 will reduce shortfalls for low-income households, three-person households, and households with four or more persons in MDOT model areas and will reduce shortfalls for zero-vehicle households, households with fewer vehicles than workers, and households with three or more persons in SEMCOG model areas.

Table 11 compares Option 1 and Option 2 (Final) relative to the number of retrieval completes for each model area. For each option, Table 11 contains the number of completes in the spring, the target number of completes for the fall, and the spring-plus-fall total. Table 11 also indicates if Option 2 (Final) increases the oversampling factor for the model area's high density stratum.

Table 11. Target Complete by County – Two Options

Model area	Target number of completes						Increase Fall completes for Option 2 (Final)	Increase oversampling for Option 2 (Final)
	Option 1			Option 2 (Final)				
	Spring	Fall	Both	Spring	Fall	Both		
SEMCOG minus WATS	794	856	1650	794	856	1650		
SMR	669	531	1200	669	568	1237	**Yes**	
NMR	699	501	1200	699	501	1200		
Small cities	523	477	1000	523	477	1000		
GVMC	705	295	1000	705	295	1000		
TCRPC	585	215	800	585	215	800		
GCMPC	352	448	800	352	503	855	**Yes**	**Yes**
GLBR	499	301	800	499	301	800		
WATS	492	308	800	492	357	849	**Yes**	**Yes**
KATS	478	322	800	478	322	800		
WestPlan	335	315	650	335	315	650		
JACTS	311	339	650	311	506	817	**Yes**	
TwinCATS/ NATS	410	390	800	410	390	800		
MACC	393	257	650	393	406	799	**Yes**	
BCATS	316	334	650	316	437	753	**Yes**	
TVC	503	147	650	503	147	650		
E. Wayne	463	807	1270	463	807	1270		**Yes**
W. Wayne	856	904	1660	856	804	1660		**Yes**
Oakland	976	1204	2180	976	1769	2745	**Yes**	**Yes**
Macomb	544	926	1470	544	926	1470		**Yes**
Monroe	245	405	650	245	716	961	**Yes**	**Yes**
St. Clair	286	364	650	286	487	773	**Yes**	**Yes**
Livingston	355	295	650	355	610	965	**Yes**	**Yes**
Total (unduplicated)	10,995	10,085	**20,980**	10,995	11,859	**22,854**		

5.3 Adjustments to Instruments and Materials

5.3.1 Clarification of Travel Date Change Language and Directions

Only minor changes were made to the instrument between the spring and fall data collection. The only front-end improvement was to strengthen the language on the travel date assignment page. During the spring field period we received a noticeable number of calls requesting a new travel date. Almost all of these requests were for situations that are not a legitimate reason to change a travel date. Often participants would request a travel date change because they do not typically travel on that day or their travel was going to be abnormal for that day of week. Our telephone staff would explain to them that we want to collect all types of travel and the only reason to change a travel date is if the entire household will be traveling outside of the state of Michigan for that entire date.

In order to cut down on participant confusion during the fall data collection we strengthened the language around changing a travel date. We changed "I cannot record my travel on this day" to "I cannot record my travel on this day because everyone in my household will be out of the state". We also added a list of acceptable and unacceptable reasons to change a travel date on the page where a participant would actually choose a new travel date. Participants will see this list only if they have already selected "I cannot record my travel on this day because everyone in my household will be out of the state". The list displays the following text:

Reasons to choose a new travel date:

- Your entire household will be traveling OUTSIDE of the state of Michigan for the entire travel day.

Reasons to keep your assigned travel date:

- One member of your household will be out of town, but others will still be in town.
- You are not planning to go to any places that day.
- You are not planning to go to any typical places that day.

5.3.2 Clarification of Incentive Structure and Qualification Language

The language describing the incentive structure on the advance letter was identified as unclear regarding whether the incentives were paid to the household, or whether there was a per-person structure. The text on the letter was updated to specifically state that the $20 incentive was being offered to the household.

The original text read: 'Receive $20 for completing all parts of the survey.' The revised text reads: 'Your household will receive $20 for completing all parts of the survey.'

5.4 Adjustments to the Public Awareness Plan

The Fall Public Awareness Plan will utilize specific strategies, tools, and tactics developed as part of the larger Public Awareness Plan. It suggests activities that are most effective at reaching study areas that experienced a relatively low response to the spring survey, as determined by the data collection analysis.

The strategies and tactics below could all be utilized to better target specific populations or geographic areas where more participation is desired.

- Messaging:
    - Be direct with the "ask." In other words, in press releases and in further targeting outreach, plead with the public and others to help MDOT/SEMCOG reach more people in specific areas. While we are satisfied with overall results, ask people to help and to take action.
    - Timing should coincide with the fall survey invitation letter mailings.
- Media Outreach:
    - Target key ethnic demographic news outlets such as the Arab-American News for MDOT/SEMCOG to pitch stories and possibly advertise in. Inquire about reporters taking part in the survey as well, and being equipped with a GPS device. SEE RELATED UNDER *"News Media Strategies"* in Section 2 of the Public Awareness Plan.
    - Target specific local news media (TV, newspapers).
    - SEE RELATED UNDER *"News Media Strategies"* in Section 2 of the Public Awareness Plan.
- County Road Commissions:
    - Working with the County Road Commission, who has agreed to help in outreach in any way, we can target specific county road commissions where more participation in a county is desired. This could include asking them to send information (press release or other) in their mailings and newsletters and engaging them on social media channels.
- Local Governments:
    - Work with targeted localities to distribute information to their constituents.
    - Also, typically local governments have lists of civic organizations and neighborhood association contacts that could be enlisted to assist.
    - SEMCOG has an e-mail list of local government contacts, a biweekly newsletter and SEMscope newsletter.
    - Contact can be made to inquire about distributing information in newsletters, on websites, social media and other outlets.
- Community Outreach:
    - To be more effective in hard to reach population areas, we must go to them. Simple tactics that can be utilized include "pop up information posterboards" and posting fliers on bulletin boards in strategic locations.
    - MDOT/SEMCOG could also choose to be part of a local event or festival which could include hosting a staffed table with MTC III/STC15 information.
    - SEE RELATED UNDER *"Community Outreach"* in Section 5 of the Public Awareness Plan.

Appendix ZZ

Data Quality Control and Geocoding Procedures Memo

Tasks 3f/14f: Develop Data Quality Control and Geocoding Procedures

This memorandum describes the quality control plan, data checking procedures, and geocoding process, for the 2015 MI Travel Counts[1] (MTC) survey. This memorandum may be updated, based on the experiences and findings following report and data deliveries.

Westat adopts procedures to ensure that our clients are provided optimal quality services and products. It is important to plan and monitor quality from the beginning of the project and continually strive to improve it throughout. A detailed, comprehensive quality assurance (QA) / quality control (QC) plan is a critical item in our survey research toolbox. Feedback protocols that proactively monitor, probe for and identify data quality issues are important elements of a successful QA/QC plan. Westat will implement several feedback mechanisms throughout our various project areas so that potential problems can be identified and resolved promptly.

It is important to have a thorough data checking process in place. Each household in the study will be subject to multiple levels of data checking and review. This document details the process to be used in the MI Travel Counts survey and starts by presenting a brief overview of the documented procedures used in the MTC II survey and follows that with the proposed updates to the data QA/QC process.

Review Previous Project Documentation

This memo follows the structure of the MTC II manual to the extent possible. However, the methods, survey platforms and tools used to conduct the MTC are different from those used in the previous effort; therefore adjustments will be made as needed.

The main revisions include the integration of geocoding into the online instruments, daily processing of completed households through an automated checking system, the automation of speed checks into the survey's continuous data flow, and the inclusion of GPS processing and data review.

Quality Control Plan

Westat has updated the subsections of the Quality Control Plan from the 2009 Quality Control, Geocoding Process, and Data Checking Manual to reflect changes in household travel survey methods and technologies that will be employed in the MTC data collection effort. The following subsections cover the main aspects of this updated plan.

Management Elements of Quality Assurance and Specific Quality Issues to Be Addressed

Quality control, by definition, includes a well-designed sampling plan, survey instruments, and respondent materials, and a carefully executed data collection operation. The design of the MTC sampling plan, survey instruments, and respondent materials were addressed in separate task submissions. The elements of the project management plan that are considered key to quality assurance include the following:

[1] In this document MTC will refer to MTC III and STC15.

- Effective involvement of MDOT and SEMCOG staff to ensure survey outcomes meet the agency needs
- Monitoring the areas of design where biases and errors can occur
- A dedicated team of supervisors, trainers, and interviewers
- Establishing and maintaining a detailed project work schedule
- Developing and maintaining detailed data collection protocols
- Effective selection, training, and debriefing of interviewers, explaining the effort and their responsibilities
- Periodic monitoring of interviews and calls in the helpdesk accompanied by continuous supervisor monitoring and feedback in the form of supplemental training
- Electronic tracking and monitoring of interviewers' performance – dialing statistics, completed interviews, refusals, non-contacts, and average interviewer lengths
- Implementing an appropriate public information effort
- Establishing measures to protect respondents' data privacy and to assure confidentiality of survey data
- Secure processing, storage, and eventual disposal of survey data, equipment, and materials

Survey Management System

Quality control requires that each household sample element be individually tracked through its lifecycle in the survey effort from sample release to completion or to final disposition. For this purpose Westat will employ its Survey Management System (SMS) specifically developed to handle multi-mode household travel surveys. This web-based sample management system will provide the up-to-date status of each household sample element through all steps of the survey process. The system's built-in call scheduler and disposition monitoring will generate continuous information to assure that each household is receiving appropriate attention so that quick remedial action can be taken as needed. Furthermore, the SMS system is tightly integrated with the survey's secure public website (https://mitravelcounts.com), its online WEB and computer-assisted telephone interview (CATI) instruments, the automated reminder program that will send emails and text messages to households to help them along the data collection process, and the continuous automated checking system. This integration will make it possible for Westat to develop reports that are continuously updated and always reflect the current state of the survey as a whole.

Automated Data Checking and Data Processing Systems

Westat's household travel survey data collection process includes several layers of automated data checks. The first set of checks is included in the online instruments, and that is followed by automated checks performed nightly on all households that complete the retrieval interview. The final round of checks happens once data delivery files are extracted from the project's database. The section Data Checks and Coding provides additional details on the automated check processes.

Definition of a Completed Household

A completed household is one in which a household-level recruitment survey has been completed, and a retrieval survey has been completed that includes the trip data for each household member. In larger households (those with more than four total household members), the household may be considered complete if no more than one household member fails to complete the retrieval survey.

In the specific case of the GPS sub-sample a completed household is one in which all household members who have been equipped with a GPS device used it at least on the assigned travel date. In larger households (those with more than four total household members), the household may be considered complete if no more than one household member fails to comply with all of the completion requirements.

These definitions were provided in the documentation of Tasks 3c/d and 14c/d Diary and GPS Survey Methods Including Incentives.

Data Checks and Coding

Data processing and data cleaning will be conducted on an ongoing basis throughout the study. Updates will be made to variables that impact data collection during the administration of survey (e.g., the removal of a household member that was originally reported in the recruit interview, but no longer lives at the household location) and at the conclusion of data collection for data that do not impact the flow of the survey (e.g., recoding trip purpose based on "Other, specify" responses).

A series of automated edits, range checks, and consistency checks will be performed within the survey instrument, and data preparation staff will perform frequency reviews and problem resolution to monitor, correct, and update the data and online programs as needed. Automated checks will be run to evaluate the validity of reported trip data and household location data.

Logic Checks Built into Online Instruments

Logic checks are programmed into the recruit and retrieval instruments to ensure that questions are answered as accurately as possible. These include requiring that key questions be answered, even if the answer is "don't know" or "prefer not to answer," and forcing the data type (e.g., requiring a number for numeric questions, US formatted telephone numbers for phone questions, etc.).

Data range checks are employed to guarantee that the data falls within the expected range for a given question (e.g., 0-110 for AGE). Skip logic is programmed into the instruments such that questions that do not apply are not asked (e.g., participants are only asked if they have driver's license if they are at least 16 years old). Specific information on these built-in checks and skip logic features can be found in the instrument design documents. In addition, the following checks are also built into the online instruments:

- If a household member reports being employed, the number of jobs reported must be at least one
- The household respondent (person 1) cannot report that they are under 18 years of age and complete recruitment
  - Must confirm they are at least 18 years of age at the beginning of recruitment
  - Cannot enter an age or age category under 18 years of age
- If an age is reported as zero, a follow-up question confirms that the person is an infant
- Households are required to provide details on as many members as they report in their household size
  - After they are done rostering the household, they are asked if there are any other household members that they need to report
- The household must confirm their home address (or PO Box) matches the sampled address
- Number of household members on a trip is not asked of one-person households
- Households that report zero vehicles are not asked if a household vehicle was used on a trip

Westat's online travel data retrieval tool, TripBuilder Web (TBW), integrates several data consistency checks within its user interface (UI). This ensures that collected places pass basic data completeness and consistency requirements. One of these checks includes the requirement that all visited places be associated with a geocoded location. That way all records deemed complete at the end of the retrieval instrument must have all destinations geocoded. Other checks built into the TBW UI include:

- All places must have valid arrival and departure times (i.e., times must increment)
- All places after the first one must have a travel mode associated with them
- First place of the day must have an arrival time equal to 3:00 am
- Last place of the day must have a departure time equal to 2:59 am
- If the trip to a place took longer than expected (based on captured distance and minimum mode speed) a follow up question will ask for a reason
- If a participant reports a single place on their travel date a follow-up question asks for the reason why no travel was reported

In addition, the following checks are performed on the captured long distance trips:

- The start date must be before and within three months of the current date (i.e., when the survey is collected)
- Must have an end date that is the same or is after its start date
- A geocode for the destination is captured
- The three-month and twelve-month frequency questions trigger confirmation questions if reported at values higher than 30 and 100 respectively

Review of Survey Data and Reporting

Frequency reviews will be conducted at the beginning, in the middle and at the end of data collection to ensure that all data were being properly captured in the survey database. Additional reviews are also performed following any changes to the survey instrument after the start of the survey.

This process consists of extracting all the collected survey data into comma-separated text files (CSV) and loading them into the statistical package R (http://r-project.org/) for summarization. The resulting report displays information for each survey variable which includes branching logic, question text, and response frequency distributions. The frequency reviews also allow for verification of the presence of conditional questions (e.g., open-end responses when "other, specify" choices are selected), and checking the range of numeric variables.

Through the review of these frequency reports, analysts can identify potential issues with the data or data collection tools and communicate their findings to the project management team so that issues can be identified and resolved.

Automated Edit Checks

Travel data will be processed through Westat's trip processing system (TPS). TPS includes a series of consistency checks on reported travel data. When a TPS edit fails, an analyst reviews the data to determine whether adjustments to the data can be made based on information provided by another household member or if the household needs to be re-contacted to resolve the inconsistency in the data. Whether the data are updated by an analyst or an interviewer as a result of a re-contact with the household, the entire household record will be reprocessed through the TPS checks. Each case is

subjected to this process until it clears TPS without any failures. Only households successfully passing these edits will be considered as complete in the final dataset. Table 1 and Table 2 identify the main automated checks currently implemented in TPS.

Table 1. Location and Travel Data Checks

- Location is missing geocode coordinates
- Location is missing full address
- Location name text contains "Home" but is not location type 1 (Home location).
- Location type 1 (Home location) text is not "HOME"
- Location name text contains "Work" but is not location type 2 (Work location).
- Location name text contains "School" but is not location type 3 (School location).
- Consecutive places use the same location or locations with identical geocode coordinates
- Consecutive locations have identical location name
- Household locations with same coordinate do not have matching addresses
- Every person in retrieved household reports at least one place
- Travel does not begin at home or does not end at home on assigned travel day
- Travel does not begin and end at same location on assigned travel day
- Zero-trip person missing response to "NOGOWHY" variable
- Place's arrival time is earlier than previous place's departure time
- Place travel speed too fast for travel mode
- Place travel speed too slow for travel mode
- Place has a person number that does not exist
- Transit travel mode assigned to a place that is not of transit type
- Transit place does not precede or follow another transit place
- Travel day tour is not closed (i.e., travel day does not start and end at the same location)

Table 2. Shared Travel Checks

- Persons report travelling together but companion count does not match
- Persons report travelling together but more than one driver reported
- Persons report travelling together but times do not match (not within 5 minutes)
- Persons report travelling together but mode does not match (allows mixing driver and passenger)
- Persons report travelling together but locations do not match (not the same location used as the places' destinations)
- Travel mode of "passenger" but the count of members on trip < 2
- Trip companion(s) expected but missing
- Place where household members disagree on number of companions
- Trip has no "driver" travel mode assigned to any member on trip where all modes are private vehicle-based

Upcoding and Cleaning

At the conclusion of the data collection period, open-ended and ”other, specify” responses will be reviewed and upcoded or collapsed as appropriate. The upcoding of responses is the activity of recoding an open-ended response into a categorical response option (e.g., recoding “slept” to “typical home activities”). The process includes removing the “other, specify” (open-ended) text response.

In addition to coding open-end text into categorical responses, Westat will also combine or collapse other responses that were similar to each other. These responses appear in the original dataset as independent responses (“one-offs”) due to misspelling of the response or different letter spacing in the response or capitalization issues. Combining these text responses allows for more efficient analysis.

Item Non-Response

The Westat data quality team will evaluate the levels of item non-response for key variables (e.g. age, income, employment status, trip purpose, etc.) at the household and person level to ensure data completeness. Relevant variables will be compiled and calculated for review by the project management team. These variables will include the total number of items asked, the number of items answered and refused, the resultant item non-response rate, and an enumerated list of refused items by household and household member. Using these data the project management team will define thresholds to be applied to the dataset that will be used to disqualify households from delivery as complete households.

The key variables reviewed and the thresholds applied to help refine the definition of a complete household are survey specific. This process is another effort made to increase the concentration of high-quality-data households while eliminating households with high levels of item non-response.

Deliverable Generation

The data deliverable generation process begins by defining data extraction queries that join data elements collected in both recruitment and retrieval survey instruments, from the WebGeoSurvey[2] (WGS) and TBW database schemas. An initial version of these queries is generated once the instruments are programmed and tested, and updated as needed based on changes to the survey instruments.

Survey variable display and branching logic from the survey instruments are programmed into these queries and updated as needed to account for unused or miscollected data values (e.g., a survey participant updated an answer to a variable which then affected their final survey path). Primary keys are documented and added to the exported copies as needed. The design of these deliverable tables is documented in an accompanying "Data Elements" file which specifies each data variable's source, type, possible responses, and the table name in which it will be delivered. Discrepancies between this design document and what the programmer finds in actual system and raw data are discussed and resolved as needed.

To ensure delivery of a final dataset that only includes samples intended for delivery, every deliverable table that contains participant data pulls from a list of participant samples where the final disposition is set as deliverable-ready. The delivered tables reference a copy of the originally collected survey and TBW data. This allows post-processing to not burden or affect the data as it was originally collected so that differences between that and the final deliverables can be generated.

[2] WebGeoSurvey is the platform used for the WEB and CATI survey components that do not include TBW.

The extraction queries are used to generate the data deliverable tables (e.g., households, persons, places, etc.). The result of these extraction queries is then exported to an external format, generally Microsoft Access, and can be converted to other formats such as SAS, CSV, or dBase files.

Once a table is exported from the database, a frequency table is generated for every variable in that table. These frequency tables are then reviewed against the "Data Elements" specifications and survey instrument logic for discrepancies. Analysts then review these frequency tables for illogical or suspicious data values for follow-up research (e.g., a frequency of one NULL value for a variable that should not have any NULL values).

Geocoding Methods and Quality Checks

Home locations are pre-geocoded as part of the address-based sample and cannot be updated during the online interview, with the exception of PO Box sample addresses[3]. These locations, along with their geocoded coordinates, are then transferred to the retrieval instrument once the household completes the recruit interview.

The online instruments collect place data for frequently visited locations, referred to as habitual locations, in the recruitment survey, and there are integrated advanced online ("on-the-fly") geocoding and points of interest (POI) resources in our WGS platform and TBW travel data retrieval tool. These resources make it possible to geocode person-level habitual locations (i.e., work and school) as part of the recruit interview. The recruit instrument requires that a geocode is collected for all habitual locations. In addition, the long distance component of the retrieval instrument will also require that a geocode of the destination be provided; wording on the question will indicate to participants that only a city-level location is expected to be provided.

The web-based UI used in TBW to retrieve travel data allows participants to select a habitual location from a list or to enter address information using free-form text. The input fields are capable of processing place name (e.g., Starbucks), street address, or intersection information to assist the respondent in providing the address of the trip destination.

If an address is provided by the respondent it is first sent to the TBW server to be searched against the SEMCOG-provided list of point addresses. On the server side the full address is then parsed using the address standardizer built into the spatial extension (http://postgis.net) of the systems' PostgreSQL relational database (http://postgresql.org) and a query is built to search for an address point match. The metaphone fuzzy string match algorithm is used when comparing the respondent-provided street names against the database to allow for common spelling errors. If a match is found it is returned to the client and displayed first on the list of candidate matches.

Only after this initial search is done are the inputs sent to the Google Maps geocoding and Google Places Search application-programming interfaces (APIs). The address data are used to search for geocode matches, while the place name is used to search for POI locations around the provided address or map center if no address is provided. A list of potential matches is then displayed to the participant in a list box and added to the electronic map using "pushpins".

Once a result is selected the geocoded coordinates are saved using longitude and latitude values in decimal degrees (WGS84 datum) along with the address components, the place name (in case of POI

[3] If a PO Box is sampled, the physical home address will be captured in the recruitment survey.

results the place name returned by the search is saved), and a geocode type. In addition to these search capabilities, the geocoding UI also allows participants to click on the map to set the geocode coordinates, or drag and drop a previously geocoded result; in these cases a reverse-geocode lookup using the updated coordinates is performed using the Google Maps API. This ensures that address information is derived for the manually adjusted coordinates.

A geocode type flag will be included in the geocode results indicating the source of the provided coordinates. Depending on the participant's selection of candidate geocodes, this flag will contain the code "semcog_address_point" for matches to the SEMCOG point address file, or geocoding types from either the Google Maps geocoding API[4] or the Google Places Search API[5]. In addition to these, the system will also record the geocode types "map_click" and "map_drag" when the user clicks on the map to capture a coordinate or drags a geocode marker on the map.

In a post-processing step Westat will attempt to match POI geocodes to the SEMCOG point address file. This process will query the SEMCOG point data set for potential matches using the captured address and coordinate information. In situations where a match is found, the POI result candidate will have its location adjusted and address information updated to match that of the SEMCOG point data set. Westat will also include key fields from the SEMCOG point address file (addtype and idnumber) in the deliverable so that these cases can be linked back to the matched points. The geocode type flag will also be extended by adding the keyword "semcog_" in front of the existing "poi_*" geocode type to indicate cases where a POI search result was matched with and updated based on data from the SEMCOG point data set.

When returning coordinates for address information, Google APIs will automatically offset coordinates from the street centerline; in regions where Google has access to parcel data it will have the geocoded coordinates placed on the center of the parcel. Westat will automate a process that will flag geocodes that fall closer than 25 feet from the MI Geographic Framework road network for adjustment. Westat will consult with MDOT to ensure that the latest version of the road network is incorporated into this process.

Westat will perform speed checks on the collected geocodes and travel data as part of the automated review process completed through TPS. These checks will detect geocoded locations that are not consistent with the reported times and travel modes. The distance used to perform these checks is derived from shortest path routes that are computed by Google Maps and integrated into the TBW UI, and the selected travel mode is used to tell Google Maps whether to compute an auto, bike or walk route (transit modes are treated as auto for the purpose of route generation). This capability removes the need to conduct further processing on a GIS tool such as TransCAD in order to derive network distances.

Within TPS each travel mode has a maximum speed that is compared against the average trip to place speed, computed using the shortest network path distance and the travel time derived from the provided arrival and departure times (see Table 3). Places that have computed average speeds above the maximum travel mode value, or below 5 mph for motorized modes or 1 mph for non-motorized modes, are flagged for analyst review.

[4] See complete list at https://developers.google.com/maps/documentation/geocoding/#Types.
[5] Preceded by "poi_", see complete list at https://developers.google.com/places/documentation/supported_types.

Table 3. Travel Modes and Maximum Speed Values

Mode ID	Name	Max Speed (mph)
1	Walk	13
2	Bicycle	20
3	Motorcycle/moped	95
4	Auto/ van/truck (as the driver)	95
5	Auto/van/truck (as the passenger)	95
6	Carpool/vanpool	95
7	School bus	50
8	Public transit û local bus	50
9	Dial-a-ride/paratransit	50
10	Private bus or shuttle	50
11	Taxi/limo	95
12	Train/Amtrak	95
13	Detroit People Mover	50
14	Airplane	600
15	Boat/ferry	50

When a check fails, analysts will review the geocoded results and captured place name and address information and use it to attempt to identify a possible alternate match using Internet search resources such as Google, Bing, and Yahoo!. In cases where additional information is needed in order to address a data inconsistency, the household record will be sent to "research." In the research status, cases are re-contacted by a specially trained member of the interviewing team to address the inconsistencies and seek a resolution from the household member. During the research call, interviewers have the ability to perform those same speed checks to verify the quality of the recollected geocode or adjusted time and travel mode information.

Finally, all collected geocodes will be checked for completeness of address attributes (i.e., address or intersection, city, state and postcode). Whenever necessary reverse geocoding using Google Maps and Bing will be used to fill in missing address components.

GPS Processing and Checks

The GPS data collected by the participants will be imported into TBW and the UTC (Universal Time Coordinate) date and time stamps in the GPS point data will be translated to local date and time. The wearable GPS data loggers will be programmed with speed filter settings that will screen out all points with zero values for speed, and points with speeds greater than zero will be recorded using a three-second interval. The initial process of creating a sequence of places based on the input stream of GPS points can be summarized in the following steps:

- Filter out GPS data based on the assigned travel date range.
- Create first place of the day so that the display can be place-based; the arrival time of this place is pre-set to 3:00am and its departure time is set to the start time of the first identified GPS trip; the coordinates for this place are also set based on the first identified trip's origin.
- Split GPS points into trips leading to places using a 120-second dwell time criteria.

- Look for mode transitions within each detected trip (non-motorized to motorized and vice-versa) and further split trips based on these detected mode transitions.
- Determine if each identified trip can be attributed to real movement or GPS noise using heuristics-based criteria.
- Check resulting trips to places against minimum travel time and distance constraints (distance covered ≥ 100 meters and travel time ≥ 1 minute); if a trip fails to meet these its points are aggregated onto the next place.
- Compare each place's destination point against the household's geocoded habitual locations and associate trip ends with location if the distance between them is less than 75 meters

Once initial GPS processing in TBW is complete, the resulting place sequences will be visually reviewed by analysts to screen out traffic delays and other falsely identified stops with dwell times of 120 seconds or more, as well as to add stops with short dwell times not identified in the initial processing but with clear stop characteristics, for example, a brief stop (under 120 seconds) at a post office / mailbox, or a fast food drive-thru. These data will be pushed to the TBW and the participating household contacted to complete the prompted recall survey online (or by telephone).

Following the prompted recall survey completion, the GPS-derived places for all other days will be processed and reviewed according to the following criteria:

- Review places with durations greater than 60 minutes or shorter than 2 minutes
- Review places with distances greater than 60 miles or shorter than 0.25 miles
- Ensure that travel mode is present for each place greater than place 1
- Trip passes all speed checks
- Maximum point speed is not greater than 85 mph
- Average speed is not greater than maximum point speed
- Review places with a gap distance greater than 0.5 miles (gap distance is the distance between the previous destination and the first GPS point on the trip to the current destination)
- Review GPS places that are outside the travel date range
- Review last GPS place of each day if it does not end at participant's home

Appendix AAA

Final Weekly Report

Westat®

An Employee-Owned Research Corporation®

Date: March 11, 2016

To: Karen Faussett, Project Managers
Michigan Department of Transportation (MDOT)

From: Jesse Casas, Martha Wilaby, Jeremy Wilhelm, and Shawn McCloskey

Subject: MI Travel Counts III: Weekly Project Progress Report – Post Deliverable

Project Overview:

This version of the weekly report includes only households flagged and included in the final dataset to be delivered to MDOT. The 24,338 recruited households remained unchanged from the final weekly report delivered on 1/8/2016.

The pre-final numbers of 8,621 retrieved households in the fall effort and 16,477 ended up at 8,420 fall households and 16,276 total households. There were originally 1,368 GPS households reported in the final weekly reported. The deliverable ended up containing 1,324.

The following is a list of tables and figures provided in this report.

Progress Report Tables:

Table 1: Participation Rates by Sample Area

Sample Area	Addresses Sampled[1]	Released	PND	Recruited Households				Retrieved Households				
				Urban	Non-urban	Total	Recruitment Rate	Urban	Non-urban	Total	Retrieval Rate	Average Trips/HH
Statewide Model												
Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County (WATS)	44,361	38,125	860			2,876	7.5%			1,910	66.4%	11.0
Southern Michigan Rural	23,858	20,284	598			1,984	9.8%			1,339	67.5%	10.7
Northern Michigan Rural	20,862	17,133	962			1,882	11.0%			1,255	66.7%	10.1
Small Cities	19,399	16,221	647			1,705	10.5%			1,124	65.9%	11.3
Statewide Model Total	108,480	91,763	2,847			8,447	9.2%			5,628	66.6%	10.8
Urban Model Areas												
Grand Valley Metropolitan Council (GVMC)	16,200	12,981	345	1,281	311	1,592	12.3%	850	216	1,066	67.0%	11.6
Tri-County Regional Planning Commission (TCRPC)	12,560	10,251	443	799	373	1,172	11.4%	570	268	838	71.5%	10.4
Genesee County Metropolitan Planning Commission (GCMPC)	20,250	17,561	406	975	560	1,535	8.7%	606	402	1,008	65.7%	11.0
Great Lakes Bay Region (GLBR)	14,588	12,045	395	769	552	1,321	11.0%	520	376	896	67.8%	10.8
Washtenaw Area Transportation Study (WATS)	15,095	12,590	334	917	393	1,310	10.4%	661	277	938	71.6%	11.2
Kalamazoo Area Transportation Study (KATS)	14,727	12,139	465	924	413	1,337	11.0%	593	284	877	65.6%	10.8
West Michigan Metropolitan Transportation Planning Program (WestPlan)	13,456	11,365	415	745	414	1,159	10.2%	468	280	748	64.5%	10.7
Jackson Area Comprehensive Transportation Study (JACTS)	17,355	15,279	486	643	799	1,442	9.4%	403	549	952	66.0%	10.8
Twin Cities Area Transportation Study (TwinCATS) and Nilles/Buchanan/Cass Area Transportation Study (NATS)	16,798	14,188	439	891	505	1,396	9.8%	552	361	913	65.4%	11.0
Macatawa Area Coordinating Council (MACC)	13,885	11,803	258	870	469	1,339	11.3%	569	339	908	67.8%	11.3
Battle Creek Area Transportation Study (BCATS)	16,401	14,519	360	809	481	1,290	8.9%	519	322	841	65.2%	10.4
Traverse City (TVC)	9,642	7,777	281	589	409	998	12.8%	388	275	663	66.4%	10.6
Urban Model Areas Total	180,957	152,498	4,627	10,212	5,679	15,891	10.4%	6,699	3,949	10,648	67.0%	10.9
Combined Total	289,437	244,261	7,474	10,212	5,679	24,338	10.0%	6,699	3,949	16,276	66.9%	10.9

[1] The definitions for columns 2-4 in Table 1 are as follow:

1. Addresses Sampled: The total addresses sampled in the sample draw, including reserve sample.
2. Released: The count of addresses that were sent an invitation
3. PND: Postal Non-Deliverable are cases that the post office returned as undeliverable. This is typically due to a vacancy.

Statewide Results

In Table 2a, the income as reported in the recruitment survey is presented. Table 2b presents the results of the income follow-up question presented in the retrieval survey and categorizes all income responses into broader bins. Of the 1,880 households that responded with "Refused" or "Don't know" to the income question the first time, 985 provided an answer in the second step, reducing the non-response rate from 11.55% to 5.50%.

Table 2a: Reported Income (Recruitment Survey)

	Recruited Households		Retrieved Households	
	Actual	% of Total	Actual	% of Total
Less than $15,000	2,451	10.07%	1,481	9.10%
$15,000 - $ 24,999	2,326	9.56%	1,528	9.39%
$25,000 - $34,999	2,424	9.96%	1,636	10.05%
$35,000 - $49,999	3,253	13.37%	2,234	13.73%
$50,000 - $74,999	4,445	18.26%	3,113	19.13%
$75,000 - $99,999	2,724	11.19%	1,890	11.61%
$100,000 - $124,999	1,803	7.41%	1,208	7.42%
$125,000 - $149,999	812	3.34%	568	3.49%
$150,000 or more	1,113	4.57%	738	4.53%
Refused	2,646	10.87%	1,694	10.41%
Don't know	341	1.40%	186	1.14%
Totals	24,338	100%	16,276	100%

Table 2b: Reported Income Follow-up (Combined)

	Retrieved Households	
Income	Actual	% of Total
Less than $25,000	3,207	19.70%
$25,000 - $49,999	4,138	25.42%
$50,000 - $74,999	3,316	20.37%
$75,000+	4,720	29.00%
Refused	826	5.07%
Don't know	69	0.42%
Totals	16,276	100.00%

Note that Table 3 has been updated to include non-responses. Since there are no targets for these, the percent of Target columns have been left blank.

Table 3: Household Income by Household Size

Household Income	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	2,010	1,678	83.5%	912	921	101.0%	380	331	87.1%	399	277	69.4%	3,701	3,207	86.7%
$25,000 - $49,999	1,198	1,343	112.1%	1,522	1,838	120.8%	487	451	92.6%	626	506	80.8%	3,833	4,138	108.0%
$50,000 - $74,999	464	648	139.7%	1,102	1,740	157.9%	469	419	89.3%	684	509	74.4%	2,719	3,316	122.0%
$75,000+	273	436	159.7%	1,450	2,250	155.2%	767	817	106.5%	1,357	1,217	89.7%	3,847	4,720	122.7%
Subtotal	**3,945**	**4,105**	**104.1%**	**4,986**	**6,749**	**135.4%**	**2,103**	**2,018**	**96.0%**	**3,066**	**2,509**	**81.8%**	**14,100**	**15,381**	**109.1%**
Refused		167	4.1%		431	6.4%		114	5.6%		114	4.5%		826	5.4%
Don't know		16	0.4%		30	0.4%		6	0.3%		17	0.7%		69	0.4%
Totals	3,945	4,288	108.7%	4,986	7,210	144.6%	2,103	2,138	101.7%	3,066	2,640	86.1%	14,100	16,276	115.4%

Table 4: Number of Household Workers

	Recruited Households		Retrieved Households		2008-2012 ACS	
Household Workers	Actual	% of Total	Actual	% of Total	Count	% of Population
0	7,015	28.82%	4,920	30.23%	1,219,417	31.93%
1	8,613	35.39%	5,916	36.35%	1,438,710	37.67%
2	7,408	30.44%	4,769	29.30%	967,739	25.34%
3+	1,302	5.35%	671	4.12%	193,065	5.06%
Total	24,338	100.00%	16,276	100.00%	3,818,931	100.00%

Table 5: Number of Available Household Vehicles

	Recruited Households		Retrieved Households		2008-2012 ACS	
Household Vehicles	Actual	% of Total	Actual	% of Total	Count	% of Population
0	1,002	4.12%	602	3.70%	292,699	7.66%
1	6,983	28.69%	4,936	30.33%	1,334,098	34.93%
2	9,896	40.66%	6,699	41.16%	1,484,159	38.86%
3	4,010	16.48%	2,575	15.82%	504,563	13.21%
4+	2,447	10.05%	1,464	8.99%	203,412	5.33%
Total	24,338	100.00%	16,276	100.00%	3,818,931	100.00%

Table 6a: Trips per Household

Trips	Households	Percent
0	417	2.56%
1	68	0.42%
2	1,102	6.77%
3	794	4.88%
4	1,681	10.33%
5	1,135	6.97%
6	1,531	9.41%
7	958	5.89%
8	1,311	8.05%
9	807	4.96%
10	988	6.07%
11	573	3.52%
12	723	4.44%
13	424	2.61%
14	494	3.04%
15	315	1.94%
16	345	2.12%
17	226	1.39%
18+	2,384	14.65%
Total	16,276	100.00%

Table 6b: Average Trips per Person by Household Count

Trips Per Person	Households	Percent
0	417	2.56%
0.1 to .9	157	0.96%
1 to 1.9	1,077	6.62%
2 to 2.9	3,003	18.45%
3 to 3.9	2,808	17.25%
4 to 4.9	2,663	16.36%
5 to 5.9	1,861	11.43%
6 to 6.9	1,277	7.85%
7 to 7.9	781	4.80%
8+	2,232	13.71%
Total	16,276	100.00%

Table 7: Student Status

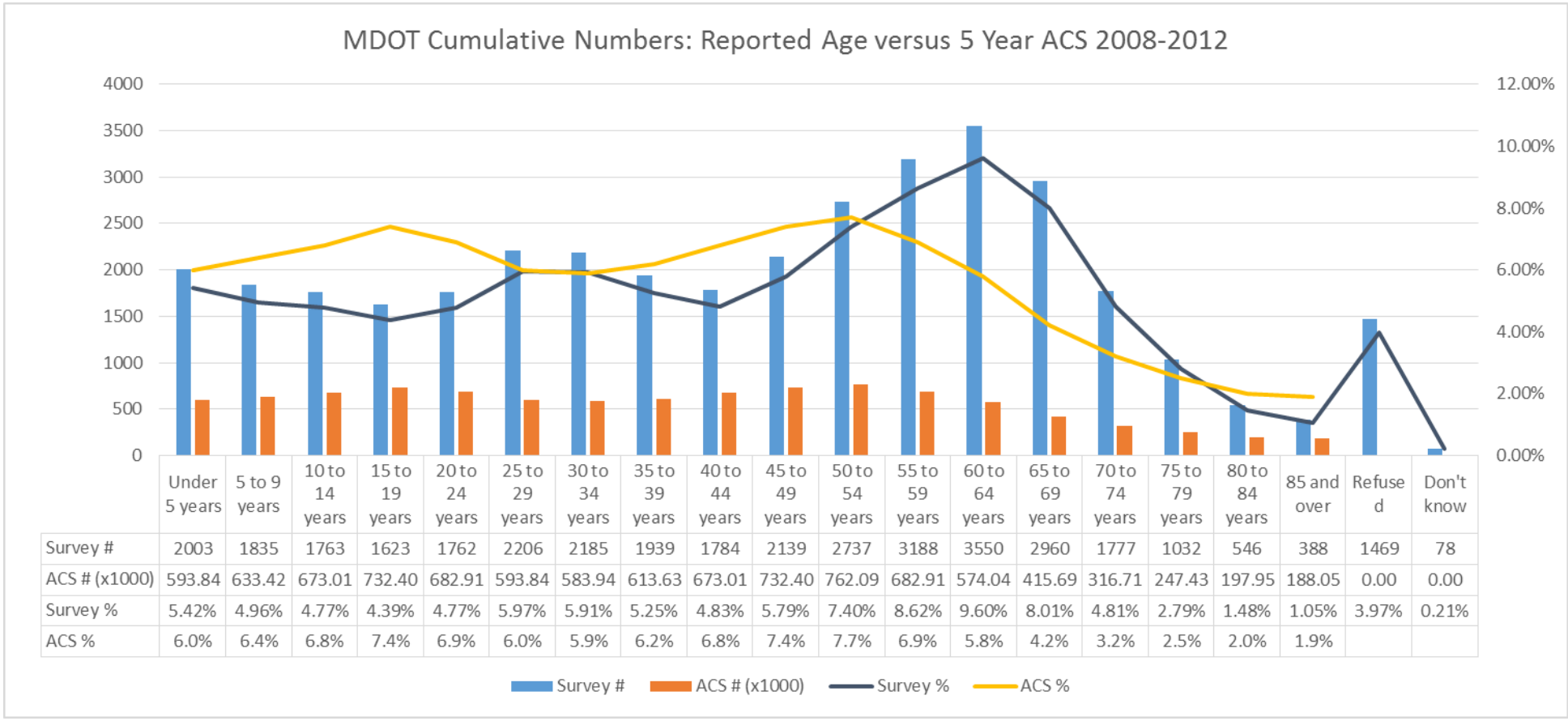
	Count of Persons	% of Persons
Full-time	6,391	17.29%
Part-time	1,274	3.45%
Not a student	28,990	78.43%
Don't know	238	0.64%
Refused	71	0.19%
Total	36,964	100.00%

Table 8: Reported Age Category

Age Category	Count of Persons	% of Persons
0-4 years old	2,078	5.62%
5-15 years old	4,180	11.31%
16-17 years old	726	1.96%
18-64 years old	22,928	62.03%
65-75 years old	5,197	14.06%
76 years old or older	1,773	4.80%
Don't know	79	0.21%
Refused	3	0.01%
Total	36,964	100.00%

Figure 1 shows responses to the initial age question and compares the results to ACS figures. Age bins collected in the age follow-up cannot be aggregated to compare to ACS (nor vice versa) because the bins were delineated for survey administration purposes (e.g., determining whether to ask employment or driver's license questions).

Figure 1: Reported Age versus 5 Year ACS 2008-2012



	Under 5 years	5 to 9 years	10 to 14 years	15 to 19 years	20 to 24 years	25 to 29 years	30 to 34 years	35 to 39 years	40 to 44 years	45 to 49 years	50 to 54 years	55 to 59 years	60 to 64 years	65 to 69 years	70 to 74 years	75 to 79 years	80 to 84 years	85 and over	Refused	Don't know
Survey #	2003	1835	1763	1623	1762	2206	2185	1939	1784	2139	2737	3188	3550	2960	1777	1032	546	388	1469	78
ACS # (x1000)	593.84	633.42	673.01	732.40	682.91	593.84	583.94	613.63	673.01	732.40	762.09	682.91	574.04	415.69	316.71	247.43	197.95	188.05	0.00	0.00
Survey %	5.42%	4.96%	4.77%	4.39%	4.77%	5.97%	5.91%	5.25%	4.83%	5.79%	7.40%	8.62%	9.60%	8.01%	4.81%	2.79%	1.48%	1.05%	3.97%	0.21%
ACS %	6.0%	6.4%	6.8%	7.4%	6.9%	6.0%	5.9%	6.2%	6.8%	7.4%	7.7%	6.9%	5.8%	4.2%	3.2%	2.5%	2.0%	1.9%		

Regional Household Income by Household Size[2]

Table 9: Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County (WATS)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	214	193	90.2%	84	83	98.8%	40	39	97.5%	50	22	44.0%	389	337	86.6%
$25,000 - $49,999	148	171	115.5%	137	141	102.9%	50	37	74.0%	64	32	50.0%	399	381	95.5%
$50,000 - $74,999	71	103	145.1%	109	162	148.6%	50	43	86.0%	68	41	60.3%	299	349	116.7%
$75,000+	52	86	165.4%	192	300	156.3%	115	145	126.1%	203	203	100.0%	563	734	130.4%
Totals	485	553	114.0%	522	686	131.4%	255	264	103.5%	385	298	77.4%	1,650	1,801	109.2%

Table 10: Southern Michigan Rural

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	146	103	70.5%	83	72	86.7%	29	28	96.6%	35	36	102.9%	293	239	81.6%
$25,000 - $49,999	86	74	86.0%	155	196	126.5%	45	52	115.6%	67	63	94.0%	353	385	109.1%
$50,000 - $74,999	28	33	117.9%	111	182	164.0%	45	41	91.1%	73	47	64.4%	257	303	117.9%
$75,000+	15	27	180.0%	115	170	147.8%	59	56	94.9%	109	80	73.4%	297	333	112.1%
Totals	275	237	86.2%	464	620	133.6%	178	177	99.4%	284	226	79.6%	1,200	1,260	105.0%

Table 11: Northern Michigan Rural

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	191	103	53.9%	100	93	93.0%	29	36	124.1%	32	28	87.5%	352	260	73.9%
$25,000 - $49,999	91	110	120.9%	183	221	120.8%	45	56	124.4%	59	47	79.7%	378	434	114.8%
$50,000 - $74,999	28	41	146.4%	114	145	127.2%	39	25	64.1%	54	40	74.1%	236	251	106.4%
$75,000+	13	18	138.5%	105	156	148.6%	44	40	90.9%	72	37	51.4%	234	251	107.3%
Totals	323	272	84.2%	502	615	122.5%	157	157	100.0%	217	152	70.0%	1,200	1,196	99.7%

[2] These tables do not include counts for the 31 households where the response to both income questions was 'Don't know' or "Refused".

Table 12: Small Cities

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	217	196	90.3%	91	106	116.5%	36	35	97.2%	38	22	57.9%	382	359	94.0%
$25,000 - $49,999	91	86	94.5%	105	149	141.9%	39	44	112.8%	46	44	95.7%	281	323	114.9%
$50,000 - $74,999	28	39	139.3%	69	107	155.1%	33	27	81.8%	40	38	95.0%	170	211	124.1%
$75,000+	10	23	230.0%	64	102	159.4%	37	21	56.8%	56	40	71.4%	167	186	111.4%
Totals	346	344	99.4%	329	464	141.0%	145	127	87.6%	180	144	80.0%	1,000	1,079	107.9%

Table 13: Grand Valley Metropolitan Council (GVMC)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	117	82	70.1%	53	48	90.6%	25	13	52.0%	30	14	46.7%	225	157	69.8%
$25,000 - $49,999	85	89	104.7%	97	107	110.3%	35	30	85.7%	49	33	67.3%	266	259	97.4%
$50,000 - $74,999	33	39	118.2%	74	123	166.2%	35	27	77.1%	61	43	70.5%	203	232	114.3%
$75,000+	17	34	200.0%	105	143	136.2%	61	58	95.1%	123	124	100.8%	306	359	117.3%
Totals	252	244	96.8%	329	421	128.0%	156	128	82.1%	263	214	81.4%	1,000	1,007	100.7%

Table 14: Tri-County Regional Planning Commission (TCRPC)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	106	75	70.8%	48	50	104.2%	24	8	33.3%	19	11	57.9%	197	144	73.1%
$25,000 - $49,999	72	83	115.3%	73	76	104.1%	25	14	56.0%	30	20	66.7%	200	193	96.5%
$50,000 - $74,999	38	55	144.7%	61	88	144.3%	26	26	100.0%	37	24	64.9%	162	193	119.1%
$75,000+	18	21	116.7%	95	135	142.1%	49	45	91.8%	79	55	69.6%	241	256	106.2%
Totals	234	234	100.0%	277	349	126.0%	124	93	75.0%	165	110	66.7%	800	786	98.3%

Table 15: Genesee County Metropolitan Planning Commission (GCMPC)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	115	140	121.7%	51	72	141.2%	29	22	75.9%	29	31	106.9%	224	265	118.3%
$25,000 - $49,999	77	78	101.3%	85	121	142.4%	28	28	100.0%	35	36	102.9%	225	263	116.9%
$50,000 - $74,999	29	32	110.3%	59	105	178.0%	27	26	96.3%	35	35	100.0%	150	198	132.0%
$75,000+	16	17	106.3%	72	118	163.9%	43	35	81.4%	70	71	101.4%	201	241	119.9%
Totals	237	267	112.7%	267	416	155.8%	127	111	87.4%	169	173	102.4%	800	967	120.9%

Table 16: Great Lakes Bay Region (GLBR)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	116	102	87.9%	55	47	85.5%	25	21	84.0%	21	15	71.4%	217	185	85.3%
$25,000 - $49,999	66	76	115.2%	95	115	121.1%	31	19	61.3%	31	27	87.1%	223	237	106.3%
$50,000 - $74,999	26	38	146.2%	63	95	150.8%	27	17	63.0%	37	17	45.9%	153	167	109.2%
$75,000+	13	20	153.8%	76	117	153.9%	42	48	114.3%	76	63	82.9%	207	248	119.8%
Totals	221	236	106.8%	289	374	129.4%	125	105	84.0%	165	122	73.9%	800	837	104.6%

Table 17: Washtenaw Area Transportation Study (WATS)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	96	61	63.5%	42	34	81.0%	16	14	87.5%	15	13	86.7%	169	122	72.2%
$25,000 - $49,999	75	69	92.0%	61	65	106.6%	18	11	61.1%	20	19	95.0%	174	164	94.3%
$50,000 - $74,999	38	65	171.1%	55	89	161.8%	21	23	109.5%	26	14	53.8%	140	191	136.4%
$75,000+	34	60	176.5%	117	173	147.9%	67	64	95.5%	100	108	108.0%	318	405	127.4%
Totals	243	255	104.9%	275	361	131.3%	122	112	91.8%	161	154	95.7%	801	882	110.1%

Table 18: Kalamazoo Area Transportation Study (KATS)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	121	92	76.0%	54	36	66.7%	24	18	75.0%	24	14	58.3%	223	160	71.7%
$25,000 - $49,999	73	64	87.7%	78	91	116.7%	26	25	96.2%	30	23	76.7%	207	203	98.1%
$50,000 - $74,999	25	28	112.0%	58	92	158.6%	26	28	107.7%	34	20	58.8%	143	168	117.5%
$75,000+	17	26	152.9%	85	124	145.9%	46	50	108.7%	79	94	119.0%	227	294	129.5%
Totals	236	210	89.0%	275	343	124.7%	122	121	99.2%	167	151	90.4%	800	825	103.1%

Table 19: West Michigan Metropolitan Transportation Planning Program (WestPlan)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	96	90	93.8%	47	48	102.1%	19	22	115.8%	20	14	70.0%	182	174	95.6%
$25,000 - $49,999	51	47	92.2%	75	83	110.7%	27	19	70.4%	32	26	81.3%	185	175	94.6%
$50,000 - $74,999	17	19	111.8%	51	83	162.7%	24	24	100.0%	35	20	57.1%	127	146	115.0%
$75,000+	8	18	225.0%	61	95	155.7%	29	33	113.8%	58	49	84.5%	156	195	125.0%
Totals	172	174	101.2%	234	309	132.1%	99	98	99.0%	145	109	75.2%	650	690	106.2%

Table 20: Jackson Area Comprehensive Transportation Study (JACTS)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	92	106	115.2%	40	56	140.0%	16	20	125.0%	24	13	54.2%	172	195	113.4%
$25,000 - $49,999	50	92	184.0%	75	111	148.0%	24	32	133.3%	27	42	155.6%	176	277	157.4%
$50,000 - $74,999	18	29	161.1%	56	100	178.6%	23	26	113.0%	34	37	108.8%	131	192	146.6%
$75,000+	12	18	150.0%	64	117	182.8%	34	35	102.9%	61	65	106.6%	171	235	137.4%
Totals	172	245	142.4%	235	384	163.4%	97	113	116.5%	146	157	107.5%	650	899	138.3%

Table 21: Twin Cities Area Transportation Study (TwinCATS) and Nilles/Buchanan/Cass Area Transportation Study (NATS)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	135	103	76.3%	60	76	126.7%	22	17	77.3%	22	13	59.1%	239	209	87.4%
$25,000 - $49,999	64	84	131.3%	82	105	128.0%	27	22	81.5%	36	17	47.2%	209	228	109.1%
$50,000 - $74,999	20	26	130.0%	59	89	150.8%	23	16	69.6%	41	20	48.8%	143	151	105.6%
$75,000+	13	16	123.1%	85	139	163.5%	40	47	117.5%	71	72	101.4%	209	274	131.1%
Totals	232	229	98.7%	286	409	143.0%	112	102	91.1%	170	122	71.8%	800	862	107.8%

Table 22: Macatawa Area Coordinating Council (MACC)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	63	52	82.5%	31	22	71.0%	12	6	50.0%	16	10	62.5%	122	90	73.8%
$25,000 - $49,999	51	79	154.9%	69	99	143.5%	22	23	104.5%	44	29	65.9%	186	230	123.7%
$50,000 - $74,999	18	42	233.3%	51	126	247.1%	28	28	100.0%	47	57	121.3%	144	253	175.7%
$75,000+	9	16	177.8%	70	141	201.4%	35	56	160.0%	84	70	83.3%	198	283	142.9%
Totals	141	189	134.0%	221	388	175.6%	97	113	116.5%	191	166	86.9%	650	856	131.7%

Table 23: Battle Creek Area Transportation Study (BCATS)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	105	127	121.0%	41	61	148.8%	22	23	104.5%	18	12	66.7%	186	223	119.9%
$25,000 - $49,999	58	72	124.1%	72	80	111.1%	27	17	63.0%	32	38	118.8%	189	207	109.5%
$50,000 - $74,999	21	24	114.3%	48	87	181.3%	23	25	108.7%	30	24	80.0%	122	160	131.1%
$75,000+	13	18	138.5%	58	103	177.6%	30	52	173.3%	52	45	86.5%	153	218	142.5%
Totals	197	241	122.3%	219	331	151.1%	102	117	114.7%	132	119	90.2%	650	808	124.3%

Table 24: Traverse City (TVC)

	Household Size														
	1 Person HH			2 Person HH			3 Person HH			4+ Person HH			Totals		
Household Income	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target	Target	Actual	% of Target
Less than $25,000	80	53	66.3%	32	17	53.1%	12	9	75.0%	6	9	150.0%	130	88	67.7%
$25,000 - $49,999	60	69	115.0%	80	78	97.5%	18	22	122.2%	24	10	41.7%	182	179	98.4%
$50,000 - $74,999	26	35	134.6%	64	67	104.7%	19	17	89.5%	32	32	100.0%	141	151	107.1%
$75,000+	13	18	138.5%	86	117	136.0%	36	32	88.9%	64	41	64.1%	197	208	105.6%
Totals	179	175	97.8%	262	279	106.5%	85	80	94.1%	126	92	73.0%	650	626	96.3%

General Survey Results

Table 25: Weekly respondent incentives paid or pending[3]

Week	Log Households	Amount	GPS Households	Amount	All Households	All Incentives
4/28/2015	1,595	$31,900	3	$150	1,598	$32,050
5/1/2015	816	$16,320	5	$175	821	$16,495
5/8/2015	1,892	$37,840	40	$2,000	1,932	$39,840
5/11/2015	559	$11,180	18	$875	577	$12,055
5/14/2015	784	$15,680	51	$2,550	835	$18,230
5/18/2015	608	$12,160	47	$2,225	655	$14,385
5/22/2015	609	$12,180	68	$3,275	677	$15,455
6/1/2015	858	$17,160	113	$6,050	971	$23,210
6/8/2015	388	$7,760	89	$4,350	477	$12,110
6/15/2015	178	$3,560	117	$5,775	295	$9,335
6/22/2015	110	$2,200	91	$4,525	201	$6,725
6/29/2015	11	$220	50	$2,500	61	$2,720
6/30/2015	3	$60	3	$225	6	$285
7/6/2015	6	$120	12	$575	18	$695
7/9/2015	4	$80	7	$375	11	$455
7/16/2015	6	$120	2	$75	8	$195
7/23/2015	1	$20	4	$225	5	$245
7/30/2015	0	$0	1	$50	1	$50
8/7/2015	0	$0	2	$100	2	$100
8/20/2015	0	$0	3	$150	3	$150
9/10/2015	6	$120	0	$0	6	$120
9/17/2015	218	$4,360	0	$0	218	$4,360
9/25/2015	711	$14,220	11	$625	722	$14,845
9/28/2015	487	$9,740	8	$400	495	$10,140
10/1/2015	684	$13,680	15	$625	699	$14,305
10/5/2015	581	$11,620	46	$2,200	627	$13,820
10/9/2015	800	$16,000	34	$1,875	834	$17,875
10/12/2015	569	$11,380	50	$2,350	619	$13,730
10/15/2015	754	$15,080	70	$3,300	824	$18,380
10/20/2015	705	$14,100	65	$3,050	770	$17,150
10/26/2015	1,029	$20,580	90	$4,175	1,119	$24,755
10/29/2015	406	$8,120	52	$2,475	458	$10,595
11/2/2015	379	$7,580	48	$2,250	427	$9,830
11/9/2015	427	$8,540	85	$3,875	512	$12,415
11/16/2015	298	$5,960	103	$5,350	401	$11,310
11/23/2015	107	$2,140	80	$3,800	187	$5,940
12/3/2015	52	$1,040	95	$4,900	147	$5,940
12/10/2015	41	$820	53	$2,450	94	$3,270
12/17/2015	23	$460	28	$1,400	51	$1,860
12/24/2015	3	$60	1	$50	4	$110
Total	16,708	$334,160	1,660	$81,375	18,368	$415,535

[3] The first round of incentives will be processed for payment during the week of 4/27/2015.

The figures presented in table 26 and 27 for contacts by email and toll free hotline are for all sample (i.e., include sample sponsored by MDOT as well as sample sponsored by SEMCOG). These now show only fall numbers. Numbers for the pilot, spring and summer are available in the separate excel file.

Table 26: Number of 800 calls by category[4]

Phase	Week	Survey Question	Joining The Survey - Invited	Website Technical Help	Incentive Status	Opting Out	Report Travel	Returning a Research Call	Joining The Survey - Not Invited	General Feedback	Lost PIN	Total
fall	8/31/2015	114	0	0	0	0	0	0	0	0	0	114
	9/7/2015	495	10	1	1	3	0	0	0	0	0	510
	9/14/2015	1,584	24	0	0	1	2	1	0	1	0	1,613
	9/21/2015	1,970	23	1	0	1	12	3	1	1	0	2,012
	9/28/2015	2,785	8	7	0	1	1	0	0	0	0	2,802
	10/5/2015	2,742	5	2	1	2	11	0	0	0	0	2,763
	10/12/2015	2,244	18	12	7	6	24	2	0	1	1	2,315
	10/19/2015	1,822	12	8	2	4	29	7	0	0	0	1,884
	10/26/2015	1,246	1	1	1	2	4	11	0	0	0	1,266
	11/2/2015	977	0	2	1	2	2	6	0	0	0	990
	11/9/2015	766	2	2	0	0	3	2	0	0	0	775
	11/16/2015	531	0	0	0	2	1	1	0	0	0	535
	11/23/2015	234	0	0	2	0	1	4	0	0	0	241
	11/30/2015	270	0	0	0	1	2	2	0	0	0	275
	12/7/2015	163	0	0	0	0	0	1	0	0	0	164
	12/14/2015	32	0	0	0	0	0	0	0	0	0	32
	12/21/2015	125	0	0	0	0	0	0	0	0	0	125
	12/28/2015	24	0	0	0	0	0	0	0	0	0	24
	1/4/2016	23	0	0	0	0	0	0	0	0	0	23
	Total	30,466	195	45	24	53	1,287	60	2	5	7	32,144

[4] The numbers in this table were incorrect in the last weekly report, but have been corrected for this report.

Table 27: Number of emails per week by category

Phase	Week	Survey Question	Joining the Survey	Website Technical Help	General Feedback	Lost PIN Code	Use of Data	Total
fall	8/31/2015	1	1	0	0	2	0	4
	9/7/2015	7	9	1	3	13	0	33
	9/14/2015	22	18	4	6	69	1	120
	9/21/2015	26	14	6	11	123	4	184
	9/28/2015	32	15	12	9	153	3	224
	10/5/2015	20	21	5	14	150	2	212
	10/12/2015	22	10	9	10	91	2	144
	10/19/2015	12	4	9	7	87	3	122
	10/26/2015	5	2	3	9	36	1	56
	11/2/2015	5	3	0	2	34	0	44
	11/9/2015	2	1	2	1	23	1	30
	11/16/2015	5	3	0	2	12	1	23
	11/23/2015	0	0	0	1	10	0	11
	11/30/2015	2	1	1	0	7	0	11
	12/7/2015	1	2	0	0	4	0	7
	12/14/2015	0	0	0	2	0	1	3
	12/21/2015	0	0	0	0	1	0	1
	12/28/2015	1	0	0	0	0	0	1
	1/4/2016	0	0	0	1	0	0	1
	Total	282	181	115	167	1,549	29	2,323

Table 28 reports GPS progress for each MDOT sample area and contains fields for the following:

- Recruited - households that joined the survey and have been or will be sent GPS equipment (all households for the pilot fall under the 'have been sent' category at the time of this report)
- Returned - households that were sent GPS equipment and have returned it to Westat
- Released - households that have had GPS data processed and are cleared to confirm travel details
- Completed - households that have confirmed all details and completed the retrieval survey

The retrieval rate reflects the percent of recruited households that have completed each stage of the survey.

Table 28: GPS progress by sample area

Sample Area	Recruited	Returned	Released	Completed	Retrieval Rate
Statewide Model					
Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County (WATS)	262	222	222	177	68%
Southern Michigan Rural	156	144	143	106	68%
Northern Michigan Rural	155	133	132	90	58%
Small Cities	150	120	120	87	58%
Statewide Model Total	723	619	617	460	64%
Urban Model Areas					
Grand Valley Metropolitan Council (GVMC)	136	121	121	84	62%
Tri-County Regional Planning Commission (TCRPC)	92	87	87	67	73%
Genesee County Metropolitan Planning Commission (GCMPC)	168	135	135	105	63%
Great Lakes Bay Region (GLBR)	108	97	96	69	64%
Washtenaw Area Transportation Study (WATS)	131	115	114	81	62%
Kalamazoo Area Transportation Study (KATS)	110	89	89	63	57%
West Michigan Metropolitan Transportation Planning Program (WestPlan)	105	88	88	63	60%
Jackson Area Comprehensive Transportation Study (JACTS)	135	111	110	78	58%
Twin Cities Area Transportation Study (TwinCATS) and Nilles/Buchanan/Cass Area Transportation Study (NATS)	117	100	98	78	67%
Macatawa Area Coordinating Council (MACC)	102	84	83	56	55%
Battle Creek Area Transportation Study (BCATS)	116	94	94	64	55%
Traverse City (TVC)	87	78	78	56	64%
Urban Model Areas Total	1,407	1,199	1,193	864	61%
Combined Total	2,130	1,818	1,810	1,324	62%

Appendix BBB

Data Delivery Observation Table

client	category	variable	comment	Westat response
MDOT	household	DEL_DATE	In the DEL_DATE field, the households new to the final delivery have 2015-01-22, but it should be 2016-01-22.	We will correct this
MDOT	household	DELIV_DRIV	Driving for work DELIV_DRIV = 1 for each of the following persons with high trip counts: Person 3180706-1 made 55 trips, mostly TPURP = 4 (Work at non-fixed work location).	2,838 persons, or 16%, reported that they drive as a part of their daily job. Of those, 92 persons or 3%, reported more than 5 work-related trips (TPURP=4 or TPURP=5), and 24 persons or 1% report more than 10 work-related trips.
MDOT	household	FINALGFLAG	The MDOT household table has 1,325 records with the value of 1 in FINALGFLAG but only 1,318 unique households in the gpsplace table. See the table below for the 7 missing sample numbers (SAMPNO 3057418 is also in the SEMCOG sample).	Seven FINALGFLAG=1 households completed the prompted recall survey with no GPS data, either due to equipment malfunction or because they over-wrote the data during the prompted recall stage. This can happen if a household switches assigned devices and then sees their data isn't representative of their travel day.
MDOT	location	GEOCODE_TY	34 home locations have a geocode type of “sample_z9” (28 from the first 3 deliveries), but all of these are correctly placed.	Unclear if you are asking for the type to be updated or if you are okay with the results
MDOT	household	HHCHILD	MDOT household table has 9 records (none from the SEMCOG sample) where HHCHILD does not match the number of children in the person table based on AGE18 and AGE_AAGE in the person table. All but one of these discrepancies involves persons defined as children on the basis of	That is correct - HHCHILD did not include AGE18 responses.
MDOT	household	HHSIZX	The sum of HHSIZX is 1 more than the number of records in the person file; HH 3213050 has HHSIZX = 3, but only 2 persons listed in the person file.	Person 3 was removed from this household because the household reported that they moved out of the household before their travel survey date. We have adjusted HHSIZX for this household.
MDOT	household	HHWORKER	SAMPNO 2436445 (HHWORKER of 2 but only one person employed in person table) was also identified in the MDOT observations document (the second person is under 16 but has responses in EMPYTYPE and WPLACE).	2436445 should be HHWORKER = 1. This has been corrected. The responses to EMPYTYPE and WPLACE should not be there but are non-response. No important work question like WKSTAT has a response.
MDOT	longdist	LD_3MTH	Regarding LD_3MTH and LD_12MTH being blank for HH 2135813, LDTRIPNO 3 (the only blank record in these fields): This trip location is the same as that of LDTRIPNO 2 for the same household. It appears that LDTRIPNO 3 should simply be removed from the file.	This erroneous record has been removed
MDOT	longdist	LD_REAS_O	LD_REAS_O – There are two separate codes for “drop off/pick up passenger” in the codebook.	We will correct this
MDOT	place	LNGTRP_R_O	LNG_TRP_R_O – There are two separate codes for “not ascertained” shown in the codebook.	We will correct this
MDOT	household	NA	32 of 182,822 place records are outside the US and Canada. These pertain to 20 persons, only 3 of which have more than one place record: Person 2021185-2 has 7 place records in Russia.	In most cases, the participant was out of the country on their travel day and reported the out-of-country place. In some cases, the participant chose to report trips they made while out-of-country on their travel day. You can filter these from analysis based on the country variable.
MDOT	person	NA	Only 2 persons have 17 or more responses of “don’t know” or “refused” (2266230-2, 3306896-4), and only 14 others have more than 5 such responses (out of 36,891 persons).	For 2266230-2, we left them in because they responded to all trip and household-level questions and P1 responded to most person-level; we thought the HH was still useful as a result of this. For 3306896-4, this was a large hosuehold and the person with high-refusals is the fourth person and likely a child. Both households provided complete trip details for
MDOT	person	NA	Four persons of age 15 or under have responses in fields they shouldn’t: 2436445-2 (age 10) in EMPYTYPE and WPLACE 3003411-3 (age 9) in NOLIC, EMPYTYPE, WPLACE, WMODE, INDUS, and TCOFF	These erroneous records have been removed
MDOT	place	NOGOWHY	Number of records showing no travel matches the person file responses to NOGOWHY (3,982).	Good!
MDOT	location	REGION	All household locations match the assigned REGION in the household file.	NA
MDOT	place	TPURP	No records have TPURP = TPURP2.	This is good. If these two matched, then the second was cleaned and coded as 00 - nothing else.
MDOT	place	TPURP	The MDOT place table has 24 records where TPURP is -9.	In most cases, the participant is only missing TPURP for one place. Some of these could be derived by the end data users based on the location being a habitual home or work place.
MDOT	person	WKSTAT	232 persons have EMPLY = 1 (Yes) but WKSTAT is null (other work-related fields have responses). JOBS = 0 for one of these and -7 or -8 for all the others. Should WKSTAT be changed to “0” (worker) for these, seeingas 99.89% of all other EMPLY = 1 persons have WKSTAT = 0?	WKSTAT is the preferred employment status variable and is not asked of 0-15 year olds. The inconsistency between EMPLY and WKSTAT was due to not setting WKSTAT=0 for those who had non-responses to JOBS but said yes to EMPLY. We have set their WKSTAT responses to 0.
MDOT	person	WMODE_O	WMODE_O – There are two different codes for “other personal transport”, one in caps and one lowercase. They should be combined.	We will correct this

Appendix CCC

Final Sampling Plan

Westat®

Improving lives through research™

MDOT-MTC III Statewide Household Travel Survey

Final Sampling Plan

December 8, 2014

Prepared for:
Michigan Department of Transportation
Bureau of Transportation Planning
425 Ottawa Street
Lansing, MI 48909

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Table of Contents

Appendixes **Page**

Appendixes (continued)

Tables

Introduction and Background 1

This section specifies the target population for the Michigan Travel Counts III Survey (MTC III), discusses sampling concepts, and provides an overview of the steps in developing a sample design for the MTC III. Section 2 discusses geographical strata and demographic post-strata. Section 3 discusses obtaining the samples of addresses, the sampling of travel days, and the sampling for GPS data collection. Section 4 discusses those aspects of the monitoring of the data collection operation that will indicate adherence to, or departures from, the MTC III sampling plan and the need for adjustments to sampling procedures. Section 5 describes procedures for hard-to-reach households.

1.1 Specifying the Target Population for MTC III

The first step in developing a sample design is to specify the target population. The target population for the MTC III survey is all households in Michigan, where a household is defined as all the people who occupy a housing unit. A housing unit may be a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied as separate living quarters. Separate living quarters are those in which the occupants live separately from any other individuals in the building and which have direct access from outside the building or through a common hall.

The MTC III target population excludes individuals living in group quarters, where a group quarters is a place where people live or stay in group living arrangement that is owned or managed by an entity or organization providing housing and/or services for the residents. These services may include custodial or medical care and other types of assistance, and residence is commonly restricted to those receiving these services. Group quarters include such places as college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, correctional facilities, and worker dormitories. The MTC III target population also excludes individuals staying in transient housing such as camp grounds, marinas, non-profit lodgings (YMCAs, youth hostels), and commercial hotels and motels.

1.2 Use of Stratified Sampling

After one defines the target population, the second step in developing a sample design is to consider the analysis goals of the survey and how these determine the desired number of completed cases. This step also considers how one's knowledge of respondent characteristics also determines elements of the sample design. The MTC III survey data will be used to develop transportation planning models for 16 sub-state areas. If a single statewide sample were to be selected for MTC III by using simple random sampling, the sample sizes in the less populous planning areas might be too small to estimate the area's model parameters with the desired degree of precision. However, the use of stratified sampling, where the first level of stratification corresponds to the 16 areas for which transportation models will be developed, permits the size of the sample to be specified for each transportation model area. Thus, the sampled sizes for less populous areas can be increased by sampling them at a higher rate relative to the sampling rates for more populous areas.

Within each transportation model area, some types of households will have lower response rates[1] than other types of households. For example, households with four or more persons have a higher response burden than do smaller households, and it is our experience that they have lower response rates. By creating secondary strata based on the prevalence of demographic characteristics associated with low response rates, one can direct additional data collection efforts to sub-areas where it is needed to compensate for low response rates and away from sub-areas where it is not needed. We refer to secondary strata that have either a high or a low prevalence of characteristics correlated with low response rates as *high density strata* and *low density strata*, respectively.

When different strata have different sampling rates or different response rates, it is necessary to use weights[2] when analyzing the collected data. If this is not done, then areas and sub-areas with higher sampling rates or higher response rates will be overrepresented relative to areas and sub-areas with lower rates. Large variations in the product of the sampling rate and the response rate across strata can result in loss of precision in estimates that aggregate data across strata. For the case of secondary strata that are based on the prevalence of demographic characteristics that are correlated with response rates, one should strive to have the oversampling factors be approximately equal to reciprocals of relative response rates. For example, if for a high density stratum it is expected that the response rate will be only 2 percent and in the corresponding low density stratum it is expected

[1] The sampling units for MTC III will be residential addresses. Consequently, the response rate for a particular MTC III area will be the area's number of completed cases at the end of data collection divided by the number of households in the area mailed survey invitations that were not returned as undeliverable. During a survey's planning phase, it is necessary to predict response rates based on one's knowledge of response rates in similar surveys.

[2] Also referred to as *household expansion factors*.

that the response rate will be 4 percent, then the oversampling factor for the high density stratum should be to 4/2 = 2. Then, if the expected ratio of the stratum response rates is realized, the product of the sampling rate and the response rate across these two strata will be equal, and there will be no loss in precision due to variation in the product of the sampling rate and the response rate.

1.3 Allocation to Sampling Strata and Post-strata

Sampling strata are sub-populations to which population elements are assigned prior to selecting a sample. This requires that the information needed to assign a population element to a sampling stratum is available during the survey's planning phase. An example of the assignment of population elements to sampling strata is the assignment of household addresses to categories based on linkages to geographical entities such as counties, townships, or Census geography.

Post-strata on the other hand are sub-populations to which completed cases are assigned following data collection, using data collected by the survey. An example of assignment to Post-strata is the assignment of completed household-survey cases based on household size determined by the survey.

The third step in developing a sample design is referred to as *allocation*, in which the sample designer distributes to defined strata or post-strata either the completed cases—i.e., cases for which data collection is successful—or the fielded cases—i.e., the cases for which data collection will be attempted. One type of allocation is *proportional allocation*, in which the number of completed or fielded cases assigned to a particular stratum is proportional to its number of population elements. *Disproportional allocation*, on the other hand, assigns more cases to some strata and fewer cases to other strata compared to the number of cases assigned by proportional allocation. If the only analysis objective is to compute estimates aggregated over the entire population, proportional allocation is the most efficient sample design if response rates, data collection costs, and the variability of the survey data are approximately the same across all the sampling strata.

If, on the other hand, one also wants to compute sub-population estimates, referred to as *domain analysis*, it may be necessary to use disproportional allocation of completed cases because the smaller sub-populations may require more completed cases than would be assigned by proportional allocation. If the sub-populations can be defined in terms of the sampling strata, the disproportional allocation of completed cases would be with respect to the sampling strata–one or more sampling strata would be allocated more cases than they would receive from a proportional allocation and

other sampling strata would receive less. The sub-populations for a domain analysis may, however, be defined in terms of post-strata instead of in terms of the sampling strata, and if proportional allocation were to be used some of the post-strata would not have a large enough expected sample size to yield the precision desired for the domain analysis. When this occurs, one must examine the relationships between the sampling strata and the post-strata and disproportionately allocate the completed cases to the sampling strata in such a way that the expected number of cases that will be completed in the smaller post-strata will yield the desired level of precision for the domain analysis.

The allocation of fielded cases determines the number of cases to be fielded in each sampling stratum. This allocation is calculated by dividing a stratum's desired number of completed cases by its expected return rate, where the return rate is the product of the eligibility rate and the response rate. If different strata have different return rates, then to obtain a proportional allocation of completed cases it is necessary to allocate fielded cases disproportionately. This is illustrated by the three examples described in Table 1-1. In Example 1, the fielded cases are allocated proportionately to the two sampling strata, there is 50 percent response in both strata, and as a result the completed cases are also allocated proportionately. In Example 2, the fielded cases remain proportionately allocated but the assumed response rate in Stratum 1 is changed to 25 percent. The Example 2 result is that the completed cases are disproportionately allocated and the weights adjusted for nonresponse are different for the two strata, which will produce a loss in precision in population-level estimates compared to when the weights are the same for the two strata. In Example 3, the fielded cases are allocated disproportionately, with the fielded cases in Stratum 1 being oversampled because of its lower response rate. The Example 3 result is that the completed cases are proportionately allocated and because of the oversampling of the fielded cases in Stratum 1 the resulting weights adjusted for nonresponse are the same in the two strata.

Table 1-1. Sample allocation examples

	Example 1		Example 2		Example 3	
	Stratum 1	Stratum 2	Stratum 1	Stratum 2	Stratum 1	Stratum 2
# Population elements (N)	2,000	1,000	2,000	1,000	2,000	1,000
# Fielded cases (n)	200	100	200	100	400	100
Response rate %	50%	50%	25%	50%	25%	50%
# Completed cases (r)	100	50	50	50	100	50
Weight, adjusted for nonresponse (N/r)	20	20	40	20	20	20

In Section 2 we discuss how completed cases will be disproportionately allocated across primary strata, and in Section 3 we discuss how fielded cases will be disproportionately allocated across the high- and low-density substrata.

Geographical Strata and Demographic Post-strata for MTC III 2

2.1 Sample Allocation

The primary sampling strata for the MTC III will be geographical areas defined by political boundaries, such as county boundaries or township lines. There will be 16 such sample areas, with Column 1 of Table 2-1 containing associated names. Maps of each sample area are included as Appendices A-1 through A-4. Appendices B-1 through B-5 contain the geographical descriptions for the 16 sample areas.

The MTC III data to be collected in four of the sample areas will be used to develop statewide transportation planning models, whereas the data collected in the other 12 sample areas will be used to develop models for urban areas as well as in the statewide models. In Table 2-1, the four areas for statewide models are listed in decreasing order by their number of occupied housing units according to the 2010 Census. Similarly, the 12 areas for urban modes are listed in decreasing order of the number of occupied housing units. The rightmost column of Table 2-1 contains the sampling rates, and the second from the right column of Table 2-1 contain the number of completed cases allocated to each sample area.

MDOT believes that the proposed sample size for each sample area is sufficient to calculate overall trip rates and trip-length values specific to each sample area. The sample sizes in the statewide model sample areas are sufficient to calculate trip rates and lengths for all purposes for the statewide passenger travel demand model. Data from the urban model areas may need to be combined to calculate trip rates for certain trip purposes, but there is an adequate sample size to determine overall trip rates for each individual urban model area. This will allow MDOT staff to perform data analysis after the MTC III survey is conducted, to determine if model areas can or cannot be combined relative to trip characteristics.

Table 2-1. Primary sampling strata

Sample area	Housing units (occupied)[3]	Sample size[4]	% Total housing units (occupied)
Statewide model			
Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County (WATS)[2]	1,707,565	1,650	0.10%
Southern Michigan Rural	386,208	1,200	0.31%
Northern Michigan Rural	306,995	1,200	0.39%
Small Cities	130,357	1,000	0.77%
Urban model areas			
Grand Valley Metropolitan Council (GVMC)	263,361	1,000	0.38%
Tri-County Regional Planning Commission (TCRPC)	183,589	800	0.44%
Genesee County Metropolitan Planning Commission (GCMPC)	169,202	800	0.47%
Great Lakes Bay Region (GLBR)	157,051	800	0.51%
Washtenaw Area Transportation Study (WATS)	137,193	800	0.58%
Kalamazoo Area Transportation Study (KATS)	110,760	800	0.72%
West Michigan Metropolitan Transportation Planning Program (WestPlan)	86,600	650	0.75%
Jackson Area Comprehensive Transportation Study (JACTS)	60,771	650	1.07%
Twin Cities Area Transportation Study (TwinCATS) and Nilles/Buchanan/Cass Area Transportation Study (NATS)[2]	57,322	800	1.40%
Macatawa Area Coordinating Council (MACC)	43,752	650	1.49%
Battle Creek Area Transportation Study (BCATS)	37,849	650	1.72%
Traverse City (TVC)	33,933	650	1.92%
TOTAL	**3,872,508**	**14,100**	0.36%

[1] The SEMCOG minus WATS sample size is based upon the Statewide Model needs.

[2] Combined: a minimum number of samples will be taken in each model area to ensure that specific trip length parameters can be calculated for each of the model areas.

[3] Source: 2010 Decennial Census.

[4] Source: MDOT Sample Size Determination Analysis.

MDOT's sample planning for MTC III took into consideration three key factors: 1) each urban area requires an adequate sample size to calculate trip lengths and distribution factors, 2) a minimum sample of 500 households is needed in each area to provide for proper validation of transferability,

and 3) trip rates are more easily transferable than trip lengths. Using the MTC I dataset for the Small Urban Model Area (SUMA) models and the Transportation Management Area (TMA) models, MDOT determined that the minimum sample size based on a stratified sample design using the total household trip rate (defined as Day 1, un-weighted, motorized, all trip purposes) for SUMAs was 621 and for TMAs was 516, yielding a 90 percent confidence interval half-width of +/- 5 percent. An assessment of sample size for a simple random sample based on various TMA and SUMA trip rate and trip length variables was also performed. The largest of the identified values, 650 households, was used as the base for each of the sample areas. MDOT staff also assumed that the larger the area, the more diversity there will be in trip making. Therefore, the number of households in each sample area was considered and natural breaks in the number of households per sample area were used to increase the sample size in a step-wise manner.

The overall sampling rate is 0.36 percent, with the larger sample areas having a sampling rate less than the overall rate and the smaller areas having sampling rates greater than the overall rate. The reason for this disproportional allocation of completed cases across the sample areas is so the smaller areas will have a larger sample—and thus a more precise estimate—than they would receive from a proportional allocation. The disproportional allocation increases the sampling rates for the smaller sample areas so that the number of completed cases allocated to each sample area is greater than or equal to 650.

Respondents to the MTC III report not only information about transportation events, but they also report demographic data about their households. The collected demographic data are used by transportation modelers to define sub-populations for domain analysis. Appendices C-1 through C-16 allocate the completed number of cases for each of the 16 sample area to post-strata based on household size and household income. Each of the appendices contains four tables--labeled A, B, C, and D—with Table D containing the allocation of completed cases to the post-strata within each sample area. Table D was developed by first obtaining the 2006-2010 CTPP/ACS data by household size and income (Table A in Appendices C-1 to C-16), calculating the percent distribution of households by size and income by post-stratum (Table B in Appendix C-1 to C-16), and then multiplying a post-stratum's percent distribution by the number of completed cases allocated to the sample area.

The allocation of the number of completed cases to the post-strata for each sample area produces data collection goals by type of household. While we understand that the expectation is that every cell goal will be met, we cannot guarantee that they will. That type of project management, a quota-

based approach, presents unnecessary bias to the study results and is not an approach that Westat recommends.

Different types of households have different propensities to participate in a household travel survey. Consequently, for some of the post-strata, the actual number of completed cases will exceed the specified goal, and for other types of households additional data collection effort will be needed to meet the goal. The determination of whether a goal has been achieved, however, requires that both household size and household income be non-missing in the collected survey data.

2.2 Handling Item Nonresponse for Household Income

Some respondents may consider the question about household income to be sensitive and/or difficult to answer. There are several strategies for minimizing income nonresponse, and the strategy that will be implemented for the MTC III depends on whether there are analysis uses for reported household income other than for the assignment of the completed case to the appropriate post-stratum. Many studies that require detailed income data use a 9-step income ladder starting around $10,000 or $15,000 and going up to greater than $150,000. If this level of detail is needed for analyses other than assignment to a post-stratum, we recommend that a 9-step income ladder appear on the recruitment instrument. However, if the only need for income data is for assigning a respondent to a post-stratum, then we recommend that only a four-step income ladder be asked—that is, the four categories needed for assignment to a post-stratum—be included on the recruitment instrument.

If there is no response to the income question on the recruitment instrument, then the retrieval instrument provide a second opportunity to ask about household income. If the recruitment respondent did not answer whatever income question was on the recruitment instrument—a 9-step income ladder or a 4-step income ladder—then the retrieval instrument will contain only a 4-step income-ladder question. If, on the other hand, the recruitment respondent did answer the income question, then no income question will appear on the retrieval instrument.

Finally, if all attempts to obtain income information from the respondent are unsuccessful, then household income can be imputed by obtaining from American Community Survey (ACS) five-year data the median household income for the block group containing the sampled household.

Obtaining the MTC III Samples 3

3.1 Using Residential Addresses as Sampling Units

Address-based sampling (ABS) will be used to obtain a representative sample of household addresses for each of the 16 sample areas. The sampling frame will be a database of addresses created by Marketing Systems Group (MSG) from the U.S. Postal Service's (USPS) Computerized Delivery Sequence (CDS) file. The CDS file contains information on all delivery point addresses serviced by the USPS, with the exception of general delivery (mail held at a main post office for recipients to claim within 30 days). USPS updates the CDS every 2 months. MSG performs street-level geocoding on all of the city-style addresses in the ABS sampling frame and determines for each address the Census Bureau geography fields down to the block level. For other types of addresses, such as P.O. boxes, MSG uses the centroid of the address's ZIP+4 area to determine the Census geography for the address. We will use MSG's geocoding for the initial sample selection but will re-geocode all addresses after sample selection for use in data collection operations and data delivery. For city-style addresses, the re-geocoding operation will improve the geocoding accuracy in areas, such as the SEMCOG area, where it will be possible to use point-level geocoding instead of street-level geocoding. It will not be possible to re-geocode sampled P.O. Box addresses until location information is provided by a respondent.

Geocoding of the addresses in the ABS sampling frame will permit us to not only determine the set of addresses to be sampled for each sample area, but will also permit the oversampling of addresses in areas that according to the ACS have a high concentration of hard-to-reach households, as determined by highly related demographic characteristics such as the area's prevalence of low income household or the area's prevalence of 4-person households. MSG will use reverse telephone directories to match the sampled addresses to landline telephone numbers. This will permit telephone interviewers to contact nonresponding households for which it was possible to determine a matching telephone number.

Each address on the ABS sampling frame contains several administrative codes that have been assigned by the U.S. Postal Service to manage the delivery of mail. The following codes are also used by survey designers to decide which addresses should or should not be exposed to sampling:

- **P.O. Box and Only-Way-to-Get-Mail (OWGM) Codes.** These codes indicate if a residential address is a P.O. Box, and, if it is, whether it is the only way for the residence to receive mail. In urban and suburban areas, P.O. boxes are usually not sampled because of the low prevalence of OWGM P.O. boxes. The middle column of Table 3-1 indicates the prevalence of OWGM P.O. boxes in the 16 sample areas. Except for the Northern Michigan Rural area the prevalence of OWGM P.O. boxes is very low. In the Northern Michigan Rural sample area, the area-level prevalence of OWGM P.O. boxes is 4.0 percent and is over 10 percent in Houghton, Iron, Keweenaw, and Mackinac counties. Except for the Northern Michigan Rural sample area, we plan to not expose P.O. boxes to sampling. In the Northern Michigan rural area, OWGM P.O. boxes will be sampled.

- **Vacant Address Code.** USPS indicates an address is vacant if it is unoccupied for 90 days or longer. Because new construction is considered by USPS to be vacant, but the USPS vacant address code may not be removed by the postal carrier immediately after occupancy, survey designers often sample USPS-designated vacant addresses. In national ABS surveys approximately 50 percent of survey invitations sent to addresses that USPS considers to be vacant are returned as undeliverable. The other survey invitations do yield some completed surveys, though their survey return rate is much lower than the rate for sampled addresses that USPS does not designate as vacant. The disadvantage of sampling USPS-designated vacant addresses is that because of their lower yields their use increases the average cost per completed case. Table 3-1 also indicates the prevalence rate for USPS-designated vacant addresses for the 16 sample areas. In four areas, the vacant address rate exceeds 5 percent: the SEMCOG-minus-WATS area, due to a vacant-address rate of 12 percent in Wayne County; the Small Cities area, due to a vacant-address rate of over 10 percent in the counties of Calhoun, Gogebic, and Manistee; the GCMPC area, which contains Flint, Michigan; and the BCATS area, which contains Battle Creek, Michigan. The vacant-address rate for the GLBR area is only 3.8 percent, but the vacant-address rate for Saginaw County, which is in the GLBR area is 5.2 percent. Because of the higher costs per completed case when USPS-designated vacant addresses are sampled and because it appears that the areas in Michigan that have high vacant-address rates are not areas where the USPS vacant address code is associated exclusively with new construction, we plan to exclude USPS-designated vacant addresses from sampling in the counties of Wayne, Calhoun, Gogebic, Manistee, and Saginaw and in the areas of GCMPC (i.e., Flint, MI) and BCATS (i.e., Battle Creek, MI).

- **Seasonal Address Code.** USPS designates an address as a seasonal address if mail is to be delivered to it only during a specified period of the calendar year. The rightmost column of Table 3-1 indicates that except for the Northern Michigan Rural area the prevalence of seasonal addresses is low. Table 3-2 indicates that in the Northern Michigan Rural area, the prevalence of seasonal addresses exceeds 5 percent in 22 of the area's 42 counties. These 22 counties are located near to one of the Great Lakes. Consequently, it is likely that residences with seasonal addresses in these counties, as well as other Michigan counties, are occupied only during the summer. We plan to exclude all seasonal addresses from sampling. A large proportion of the excluded addresses will be for summer-only residences. The number of excluded, seasonal addresses for households occupied in the Spring or Fall will be very small relative to the total number of non-seasonal addresses.

Table 3-1. Counts of residential addresses and prevalence of only-way-to-get-mail (OWGM) P.O. Boxes, vacant addresses, and seasonal addresses

Sample area	# Residential addresses[1]	Prevalence of OWGM P.O. Boxes[2]	Prevalence of Vacant addresses[3]	Prevalence of Seasonal addresses[3]
Statewide models				
SEMCOG minus WATS	1,893,496	0.1%	6.8%	0.0%
Southern Michigan Rural	404,022	1.1%	1.0%	1.6%
Northern Michigan Rural	317,233	4.0%	1.4%	5.7%
Small Cities	146,365	0.4%	5.8%	0.3%
Urban model areas				
GVMC	283,717	0.1%	2.3%	0.1%
TCRPC	194,321	0.2%	3.0%	0.1%
GCMPC	187,336	0.3%	9.1%	0.0%
GLBR	167,320	0.3%	3.8%	0.1%
WATS	146,537	0.3%	2.0%	0.0%
KATS	119,781	0.7%	2.6%	0.2%
WestPlan	96,205	0.2%	3.4%	0.6%
JACTS	65,407	0.7%	3.6%	0.0%
TwinCATS + NATS	62,305	0.0%	3.4%	0.7%
MACC	47,866	0.5%	1.5%	0.3%
BCATS	42,517	0.4%	5.3%	0.2%
TVC	38,161	1.1%	1.4%	1.9%
TOTAL	**4,212,589**	0.6%	4.7%	0.7%

[1] Includes vacant addresses and seasonal addresses, excludes P.O. Boxes.

[2] Denominator is number of residential addresses, including OWGM P.O. Boxes.

[3] Denominator is number of residential addresses, excluding P.O. Boxes.

- **Drop-Point-Address Code.** A drop-point address is an address to which USPS delivers mail for multiple households and then a non-USPS person distributes mail to the individual households. For example, an apartment building in which the postal carrier leaves the mail at the front desk is a drop-point address, but an apartment building in which the postal carrier puts the mail for each unit in the unit's mailbox in the lobby is not a drop-point address. The individual households associated with a drop-point address are referred to as *drop units*. We plan to sample drop units at the same rate as non-drop-point addresses. (For example, if the sampling rate is 12.5 percent, a drop point containing eight drop units will be selected one time, and one of its drop units will be sent a survey invitation. A drop point containing 16 drop units will be selected twice and two of its drop units will be sent survey invitations.) If a sampled drop-point address has a large number of drop units, we will attempt to obtain a list of the unit number for the drop units and then in a second stage of sampling randomly select specific units within the drop-point address. For the entire state of Michigan, the proportion of residences that are drop units is less than 1 percent. In the portion of Houghton County that is within the Northern Michigan area, however, 19 percent of the residences are drop units, and the prevalence of drop units for the portions of

Houghton County and Marquette County that are within the Small Cities area are 12 percent and 15 percent, respectively.

Table 3-2. Proportion of residential addresses designated as seasonal for counties in the Northern Michigan Rural sample area

County[1]	# Residential addresses[2]	Prevalence of seasonal addresses[3]
Alcona	5,574	**10.60%**
Alger	2,993	2.17%
Alpena	8,663	2.79%
Antrim	9,536	**11.60%**
Arenac	6,620	**5.50%**
Baraga	3,316	1.69%
Benzie	7,293	**12.13%**
Charlevoix	12,205	**15.39%**
Cheboygan	11,807	**11.53%**
Chippewa	8,055	**5.71%**
Clare	14,180	**6.26%**
Crawford	6,190	**7.82%**
Delta	10,816	1.81%
Dickinson	5,665	0.05%
Emmet	12,827	**6.10%**
Gladwin	12,785	**5.40%**
Gogebic	4,243	3.96%
Grand Traverse	3,368	1.13%
Houghton	9,454	1.49%
Hillsdale	13,540	**7.97%**
Iron	5,547	3.15%
Kalkaska	7,217	**7.15%**
Keweenaw	906	**28.81%**
Lake	4,808	**5.10%**
Leelanau	7,452	**7.82%**
Luce	2,542	4.21%
Mackinac	4,458	**6.42%**
Manistee	7,465	**6.11%**
Marquette	13,426	1.10%
Mason	9,701	**8.74%**
Menominee	6,155	0.36%
Missaukee	5,701	1.88%
Montmorency	4,410	4.47%
Ogemaw	9,868	3.88%
Ontonagon	2,456	2.81%
Osceola	9,083	1.60%
Oscoda	3,589	**6.88%**
Otsego	10,126	**7.04%**
Presque Isle	6,733	**7.68%**
Roscommon	13,376	4.34%
Schoolcraft	4,301	0.30%
Wexford	8,783	0.59%
TOTAL	**317,233**	5.72%

[1] Excludes parts of counties that are included in the Small Cities statewide model area or in an urban model area.

[2] Includes vacant addresses and seasonal addresses, excludes P.O. Boxes.

[3] Denominator is number of residential addresses, excluding P.O. Boxes.

3.2 Creation of Substrata for Disproportionately Allocating Fielded Cases

The cell goals for the number of completed cases specified in Table D of Appendices C-1 through C-16 are based on the distribution of household sizes and household incomes present in each sample area for the period 2006 through 2010. There will probably be some changes in these distributions between 2006 and 2010 and the time of MTC III in 2015, but we suspect these changes will be small. Consequently, if MTC III response rates are the same in each of 16 poststratification cells based on household size and household income then selecting a sufficiently large equal-probability sample of addresses in each of the 16 sample areas would yield the specified number of completed cases in each poststratification cell. The reported response rates for MTC I and MTC II, and our own experiences in conducting regional travel surveys, however, indicate that response rates differ as a function of household size and household income. Consequently, we plan to select a disproportionately stratified sample of addresses in each sample area so that more cases will be fielded in areas that have a high prevalence of hard-to-reach households.

Within each of the 16 sample areas, we will create two substrata. One of the substrata, referred to as the **high-density stratum**, will be households that, according to the 2008-2012 ACS, are located in Census block groups that have a high density of characteristics associated with hard-to-reach households. The other sampling substratum, referred to as the **low-density stratum**, will be the households in all the other block groups in the sample area. The fielded cases in the high-density stratum will be oversampled, with the amount of oversampling determined by how different the response rates are between the two substrata.

For example, Table 3-3 illustrates the creation of the two substrata for the WestPlan area, using 2008-2012 ACS data. The high-density stratum contains two substrata: one is labeled H1 and contains block groups in which the prevalence of low-income households is equal to or greater than 40 percent; the other substrata is labeled H2 and contains block groups not in substrata H1 in which the prevalence of households with 4 or more persons is at least 29 percent. The number of occupied households in the high-density stratum is 39 percent of all occupied household in the WestPlan area, but the high-density stratum contains 55 percent of the households that have an annual income of less than $25,000 and contains 47 percent of the households with 4 or more persons. Appendix D describes the substrata for each of the 16 MDOC sample areas.

Table 3-3. Substrata for the WestPlan area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# Block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	40%	34%		11%	35	17%
	H2: 4+ persons		20%	29%	36%	40	22%
	Total		55%		47%	75	39%
Low density	L: Other HHs		45%		53%	100	61%

Additionally, Table 3-4 illustrates the creation of the two substrata for the Northern Michigan Rural area. Like the substrata for the WestPlan area, the high-density stratum contains two substrata: H1, containing block groups with a high prevalence of low-income households; and H2, containing block groups with a high prevalence of households with 4 or more persons. In both Table 3-3 and Table 3-4, the number of occupied households in the high density stratum is 39 percent of all occupied household in the sample area. In the Northern Michigan Rural area the high density stratum contains 44 percent of the households that have an annual income of less than $25,000, whereas for the WestPlan area the high density stratum contained 55 percent of these households. This indicates that low income households more clustered together in the WestPlan area than in the Northern Michigan Rural area. For households with 4 or more persons, however, the North Michigan Rural area's high density stratum contains 51 percent of all the households of this type whereas the WestPlan area's high density stratum contains 47 percent. This indicates that households with 4 or more persons are more clustered together in the Northern Michigan rural area than in the WestPlan area.

Table 3-4. Substrata for the Northern Michigan rural area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	41%	23%		12%	94	14%
	H2: 4+ persons		21%	21%	39%	160	26%
	Total		44%		51%	254	39%
Low density	L: Other HHs		56%		49%	450	61%

Norman and Sigman[3] provide details on Westat's experience in using the above-described stratified-ABS sampling procedure for the 2007 National Health Interview Survey. More recently, Westat has also used this approach to sample households for travel surveys in western North Carolina, South Jersey, and Las Vegas.

[3] Norman, G. and Sigman, R. (2007). Using addresses as sampling units in the 2007 Health Information National Trends Survey, Proceedings of the Survey Research Section, American Statistical Association, Alexandria, VA.

3.3 Samples for Spring and Fall Data Collection Periods

Data will be collected in spring 2015 and in fall 2015. We plan to select the addresses for the two data collection periods independently. The two samples will be de-duplicated so that a household will be asked to participate only once in the MTC-III. . For the spring data collection period, our aim is to obtain 38 percent of the completes that have allocated to each sample area for both data collection periods. Statewide, this will be 14,100 x 0.38 = 5,358 MDOT-funded completes. Assuming 15 percent of the mailed survey invitations will be undeliverable, the overall recruitment rate will be 5 percent, and the overall retrieval rate will be 62 percent, we plan to field 5,358/[0.85 x 0.05 x 0.62] = 203,000 survey invitations for the spring data collection. These will be allocated to sample areas on the basis of the target number of completes and the MTC I participation rates for each sample area. For the spring data collection, we plan to allocate each sample area's number of fielded survey invitations to the area's high- and low-density substrata by assuming that the overall response rate in the high-density stratum will be 80 percent of the response rate in the low-density stratum.

Following the spring data collection, we will calculate the response rates for each of the 32 substrata and will identify post-stratification cells in which the number of completed cases is less than expected. This information will be used to adjust the substratum sampling rates for the sample that will be selected before the fall data collection period.

3.4 Sampling for Travel Days and GPS Data Collection

The sample of addresses selected within each sample area and substratum will be randomly partitioned into four subsamples. The addresses in each subsample will be assigned to the same travel day of week. Westat has used this method of assigning the travel day of week during the sampling stage for all of our household travel surveys. We find this approach provides random and generally smooth balance across day of week without requiring a quota sample approach. After travel day assignment, Westat will randomly select 40 percent of the sampled addresses and flag them for invitation to participate in the GPS subsample. We select a higher percentage than planned for inclusion to account for non-responding, ineligible, and opt-out households.

4 Sample Monitoring

During the two data collection periods we plan to produce weekly reports for use in managing the ongoing survey operations. The full contents of these weekly reports will be specified in the methodology report. The following weekly reports will indicate adherence to, or departures from, the sampling plan:

- The number of completed cases by sample area, substrata, and post-strata.
- Recruitment rates by sample area and substrata. The calculation of disaggregated recruitment rates and retrieval rates requires that each fielded case can be assigned to a disaggregation cell. This is possible when the disaggregation cells are the sample areas and the substrata. This is not possible for the post-strata, but we can use ACS data and the definitions of the substrata to estimate the distribution of the fielded cases across the post-strata. These estimates will be used to compute estimated recruitment rates for the post-strata.
- Retrieval rates by sample area, substrata, and post-strata.
- Achievement rates for the sample areas and post-strata, where an achievement rate is the ratio of the number of completed cases in a cell to the associated data-collection goal for the cell.

In addition to the above described weekly reports, a report will be prepared indicating for each post-strata if the data collection goal has been met and, if not, how many completes are needed during the reminder of data collection in order for the goal to be met. This report will be produced between the spring and fall data collection periods and several times near the end of the fall data collection period.

Procedures for Hard-to-Reach Populations 5

The sampling of addresses instead of phone numbers improves the ability to target some hard-to-reach populations by using ACS block-group level data to identify areas with a high concentration of such households--for example, areas with a high concentration of low income households or a high concentration of households with four or more persons. The creation of a high-density substratum within each sample area permits more cases to be fielded in areas where there is a high concentration of hard-to-reach households, which increases the recruitment effort directed toward such households. Another way that recruitment of hard-to-read households can be increased is to target the telephone followup for recruitment to households in the high-density substratum. Telephone follow-up for retrieval can also be targeted to hard-to-reach households by using information provided by the recruitment respondent.

Even if sample results fall short of the goals for some cells, this does not undermine the statistical validity of analysis results for the collected data. Generally, subgroups that fall short of the goals will have less precision than designed for, but the loss of the precision may or may not be meaningful depending on the goal for the cell, the magnitude of the shortfall, and how the data are analyzed. For example, if an additional poststratification cell has a goal of 400, but we were able to obtain only 350 completed households, there is a good chance that this shortfall of 50 completes will have a very minimal effect on whatever analysis is run. Moreover, use of survey weights offset the effects of a shortfall. Survey weights when applied to the survey data ensure that the survey estimates reflect the population since they compensate for non-response and use multiple dimensions to post-stratify (rake) to population totals.

A third component of our plan to survey hard-to-reach population groups is through a well-crafted public awareness campaign. The public awareness plan will address some of the ways in which we will attempt to improve response among the more difficult to find and to survey populations.

Appendix A1
Northern and Southern Michigan Rural

Appendix A-2
Small Cities

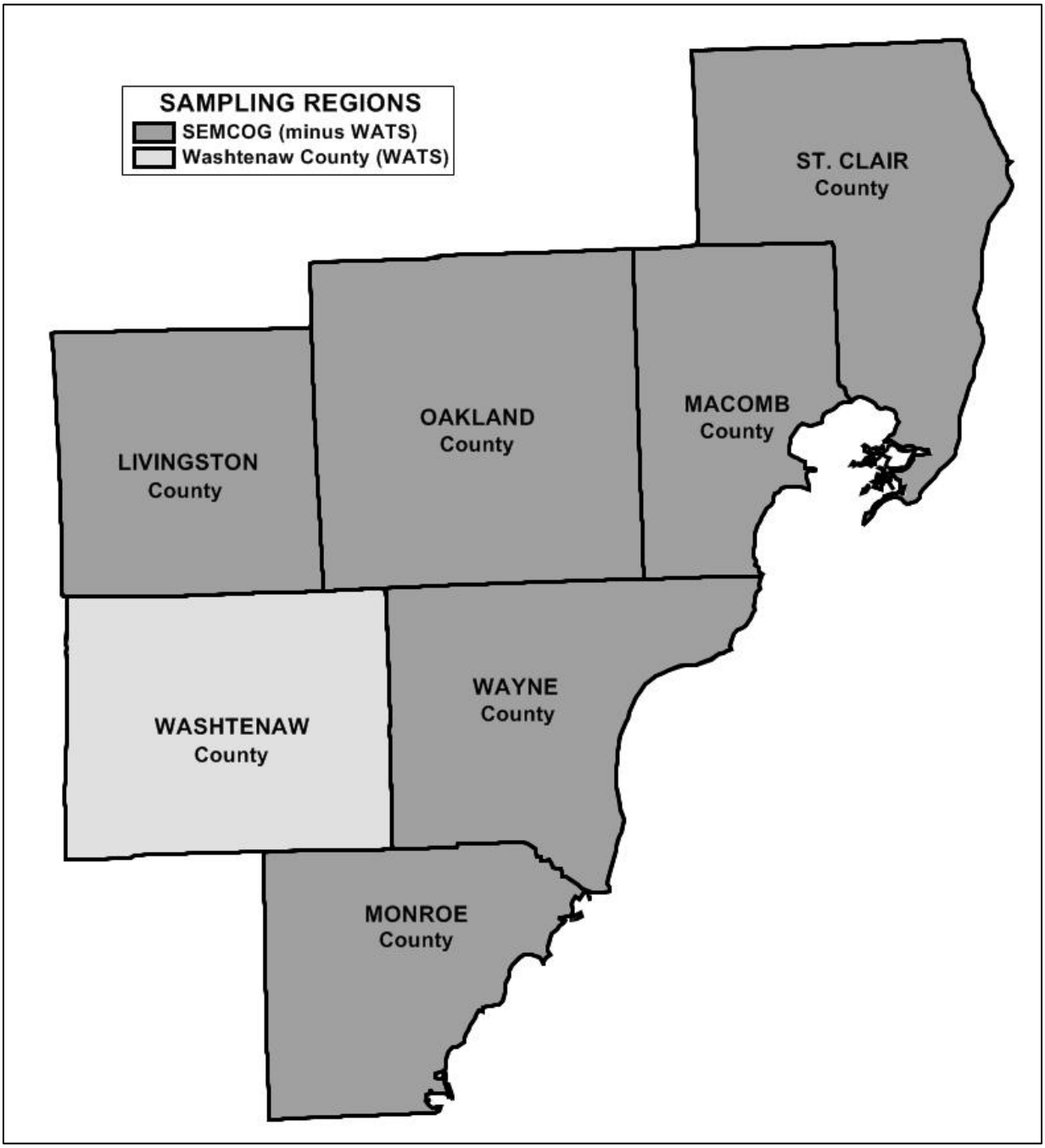
Appendix A-3
SEMCOG and Washtenaw County (WATS)

Appendix A-4
Urban Model Areas

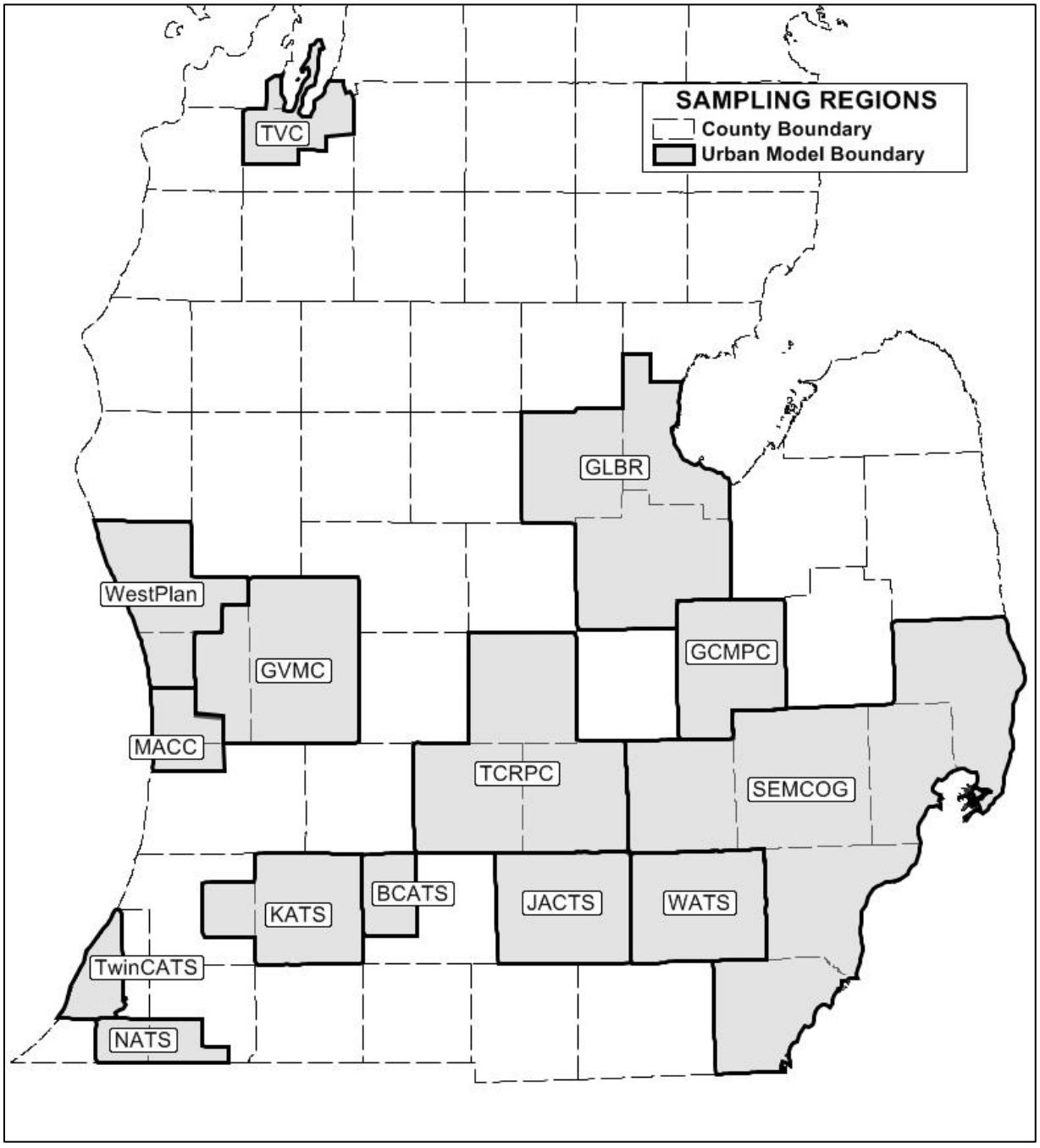
Appendix A-1
Northern and Southern Michigan Rural

Appendix A-2
Small Cities

Appendix A-3
SEMCOG and Washtenaw County (WATS)



Appendix A-4
Urban Model Areas



Appendix B-1
SEMCOG Minus Washtenaw County

Appendix B-2
Southern Michigan Rural

Appendix B-3
Northern Michigan Rural

Appendix B-4
Small Cities

Appendix B-5
TMAs and SUMAs

Appendix B-1
SEMCOG Minus Washtenaw County

Sampling Area Definition – Urban Model Area

2010 Decennial Census Housing Units

Detroit Area (Six County Region)[1]	Description	Housing Units (occupied)
Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County	Livingston County	67,380
	Macomb County	331,667
	Monroe County	58,230
	Oakland County	483,698
	St. Clair County	63,841
	Wayne County	702,749
SEMCOG Totals		**1,707,565**

[1] excludes Washtenaw County covered by the WATS TMA

Appendix B-2
Southern Michigan Rural

Sampling Area Definition – Rural Regions by County Excluding Urban Model Areas and Small Cities

2010 Decennial Census Housing Units

County	Description	Housing Units (occupied)
Allegan County	All of County, except MACC	34,935
Barry County	All of County, except Hastings City	19,641
Berrien County	All of County, except TwinCATS & NATS	14,314
Branch County	All of County, except Coldwater City	12,164
Calhoun County	All of County, except BCATS and Albion & Marshall Cities	10,152
Cass County	All of County, except NATS and Dowagiac City	9,685
Gratiot County	All of County, except Alma & St. Louis Cities	9,893
Hillsdale County	All of County, except Hillsdale City	14,822
Huron County	All of County	14,348
Ionia County	All of County, except Belding & Ionia Cities	17,555
Isabella County	All of County, except Mt. Pleasant City	17,210
Lapeer County	All of County, except Lapeer City	29,330
Lenawee County	All of County, except Adrian & Tecumseh Cities	26,079
Mecosta County	All of County, except Big Rapids City	12,771
Montcalm County	All of County, except Greenville City	19,968
Newaygo County	All of County	18,406
Oceana County	All of County	10,174
St. Joseph County	All of County, except Sturgis & Three Rivers Cities	16,108
Sanilac County	All of County	17,132
Shiawassee County	All of County, except Owosso City	21,153
Tuscola County	All of County	21,590
Van Buren County	All of County, except KATS	18,778
Southern Michigan Rural Totals		**386,208**

Appendix B-3
Northern Michigan Rural

Sampling Area Definition – Rural Regions by County Excluding Urban Model Areas and Small Cities

2010 Decennial Census Housing Units

County	Description	Housing Units (occupied)
Alcona County	All of County	5,089
Alger County	All of County	3,898
Alpena County	All of County, except Alpena City	8,057
Antrim County	All of County	9,890
Arenac County	All of County	6,701
Baraga County	All of County	3,444
Benzie County	All of County	7,298
Charlevoix County	All of County	10,882
Cheboygan County	All of County	11,133
Chippewa County	All of County, except Sault Ste. Marie City	8,334
Clare County	All of County	12,966
Crawford County	All of County	6,016
Delta County	All of County, except Escanaba City	10,370
Dickinson County	All of County, except Iron Mountain & Kingsford Cities	5,773
Emmet County	All of County, except Petoskey City	11,063
Gladwin County	All of County	10,753
Gogebic County	All of County, except Ironwood City	4,517
Grand Traverse County	All of County, except TVC	3,374
Houghton County	All of County, except Hancock & Houghton Cities	9,970
Iosco County	All of County	11,757
Iron County	All of County	5,577
Kalkaska County	All of County	6,962
Keweenaw County	All of County	1,013
Lake County	All of County	5,158
Leelanau County	All of County, except TVC	7,276
Luce County	All of County	2,412
Mackinac County	All of County	5,024
Manistee County	All of County, except Manistee City	7,492
Marquette County	All of County, except Ishpeming, Marquette & Negaunee Cities	14,453
Mason County	All of County, except Ludington City	8,391
Menominee County	All of County, except Menominee City	6,487
Missaukee County	All of County	5,843
Montmorency County	All of County	4,416
Ogemaw County	All of County	9,283
Ontonagon County	All of County	3,258
Osceola County	All of County	9,222
Oscoda County	All of County	3,772
Otsego County	All of County	9,756
Presque Isle County	All of County	5,982
Roscommon County	All of County	11,433
Schoolcraft County	All of County	3,759
Wexford County	All of County, except Cadillac City	8,741
Northern Michigan Rural Totals	**621,037**	**306,995**

Appendix B-4
Small Cities

Sampling Area Definition – Cities With Population 5,000 to 50,000 Outside Urban Model Areas
2010 Decennial Census Housing Units

Small City	County	Housing Units (occupied)
Alpena City	Alpena	4,734
Hastings City	Barry	2,910
Coldwater City	Branch	4,255
Albion City	Calhoun	2,923
Marshall City	Calhoun	3,092
Dowagiac City	Cass	2,337
Sault Ste. Marie City	Chippewa	5,995
Escanaba City	Delta	5,622
Iron Mountain City	Dickinson	3,362
Kingsford City	Dickinson	2,224
Petoskey City	Emmet	2,538
Ironwood City	Gogebic	2,520
Alma City	Gratiot	3,468
St. Louis City	Gratiot	1,491
Hillsdale City	Hillsdale	2,970
Hancock City	Houghton	1,882
Houghton City	Houghton	2,380
Belding City	Ionia	2,161
Ionia City	Ionia	2,428
Mount Pleasant City	Isabella	8,376
Lapeer City	Lapeer	3,446
Adrian City	Lenawee	7,831
Tecumseh City	Lenawee	3,604
Manistee City	Manistee	2,816
Ishpeming City	Marquette	2,824
Marquette City	Marquette	8,321
Negaunee City	Marquette	1,940
Ludington City	Mason	3,549
Big Rapids City	Mecosta	3,330
Menominee City	Menominee	3,987
Greenville City	Montcalm	3,464
Sturgis City	St. Joseph	4,088
Three Rivers City	St. Joseph	3,048
Owosso City	Shiawassee	6,161
Cadillac City	Wexford	4,280
Small City Totals		**130,357**

Appendix B-5
TMAs and SUMAs

Sampling Area Definition – Urban Model Areas

2010 Decennial Census Housing Units

Urban Model Areas	Description	Housing Units (occupied)
Grand Valley Metropolitan Council (GVMC)	Kent County; Part of Ottawa County {Allendale Charter Township, Blendon Township, Chester Township, Georgetown Charter Township, Jamestown Charter Township, Polkton Charter Township, Tallmadge Charter Township, Wright Township, Coopersville City, Hudsonville City}	263,361
Tri-County Regional Planning Commission (TCRPC)	Clinton County; Eaton County; Ingham County; Part of Shiawassee County {Part of Woodhull Township}	183,589
Genesee County Metropolitan Planning Commission (GCMPC)	Genesee County	169,202
Great Lakes Bay Region (GLBR)	Bay County; Midland County; Saginaw County	157,051
Washtenaw Area Transportation Study (WATS)	Washtenaw County	137,193
Kalamazoo Area Transportation Study (KATS)	Kalamazoo County; Part of Van Buren County {Almena Township, Antwerp Township, Paw Paw Township, & Waverly Township}	110,760
West Michigan Metropolitan Transportation Planning Program (WestPlan)	Muskegon County; Part of Ottawa County {Crockery Township, Grand Haven Charter Township, Robinson Township, Spring Lake Township, Ferrysburg City, Grand Haven City}	86,600
Jackson Area Comprehensive Transportation Study (JACTS)	Jackson County	60,771
Twin Cities Area Transportation Study (TwinCATS)[1]	Part of Berrien County {Baroda Township, Benton Charter Township, Hagar Township, Lake Charter Township, Lincoln Charter Township, Oronoko Charter Township, Royalton Township, Sodus Township, St. Joseph Charter Township, Benton Harbor City, Bridgman City, St. Joseph City}	33,981
Niles/Buchanan/Cass Area Transportation Study (NATS)[1]	Part of Berrien County {Bertrand Township, Buchanan Township, Niles Township, Buchanan City, Niles City (pt)}; Part of Cass County {Howard Township, Jefferson Township, Mason Township, Milton Township, Ontwa Township, Niles City (pt)}	23,341
Macatawa Area Coordinating Council (MACC)	Part of Allegan County {Fillmore Township, Laketown Township, Overisel Township, Holland City (pt)}; Part of Ottawa County {Holland Charter Township, Olive Township, Park Township, Port Sheldon Township, Zeeland Charter Township, Holland City (pt), Zeeland City}	43,752
Battle Creek Area Transportation Study (BCATS)	Part of Calhoun County {Bedford Charter Township, Emmett Charter Township, Leroy Township, Newton Township, Pennfield Charter Township, Battle Creek City, Springfield City}	37,849
Traverse City (TVC)	Part of Grand Traverse County {Acme Township, Blair Township, East Bay Township, Garfield Charter Township, Green Lake Township, Long Lake Township, Peninsula Township, Whitewater Township, Traverse City City (pt)}; Part of Leelanau County {Elmwood Charter Township, Traverse City City (pt)}	33,933
Urban Model Area Totals		**1,341,383**

[1]Combined: a minimum number of samples will be taken in each model area to ensure that specific trip length parameters can be calculated for each of the model areas

Appendix C-1
SEMCOG Minus Washtenaw County

Appendix C-2
Southern Michigan Rural

Appendix C-3
Northern Michigan Rural

Appendix C-4
Small Cities

Appendix C-5
GVMC

Appendix C-6
TCRPC

Appendix C-7
GCMPC

Appendix C-8
GLBR

Appendix C-9
WATS

Appendix C-10
KATS

Appendix C-11
WestPlan

Appendix C-12
JACTS

Appendix C-13a
TwinCATS

Appendix C-13b
NATS

Appendix C-13c
TwinCATS and NATS Combined

Appendix C-14
MACC

Appendix C-15
BCATS

Appendix C-16
TVC

Appendix C-1
SEMCOG Minus Washtenaw County
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	219,365	86,655	41,235	51,670	*398,925*
$25,000 – $49,999	151,925	140,425	51,445	65,680	*409,475*
$50,000 – $74,999	73,290	112,275	51,210	69,825	*306,600*
$75,000+	53,185	197,445	118,265	208,590	*577,485*
TOTAL HHS	*497,765*	*536,800*	*262,155*	*395,765*	***1,692,485***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	13.0%	5.1%	2.4%	3.1%	*23.6%*
$25,000 – $49,999	9.0%	8.3%	3.0%	3.9%	*24.2%*
$50,000 – $74,999	4.3%	6.6%	3.0%	4.1%	*18.1%*
$75,000+	3.1%	11.7%	7.0%	12.3%	*34.1%*
TOTAL HHS	*29.4%*	*31.7%*	*15.5%*	*23.4%*	***100.0%***

C: Census 2010 HHs (Based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	221,320	87,427	41,602	52,130	*402,479*
$25,000 – $49,999	153,279	141,676	51,903	66,265	*413,123*
$50,000 – $74,999	73,943	113,275	51,666	70,447	*309,332*
$75,000+	53,659	199,204	119,319	210,449	*582,630*
TOTAL HHS	*502,200*	*541,583*	*264,491*	*399,291*	***1,707,565***

D: Sample Cell Targets (Region Sample Size = 1,650)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	214	84	40	50	*389*
$25,000 – $49,999	148	137	50	64	*399*
$50,000 – $74,999	71	109	50	68	*299*
$75,000+	52	192	115	203	*563*
TOTAL HHS	*485*	*523*	*256*	*386*	***1,650***

Appendix C-2
Southern Michigan Rural
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	48,475	27,497	9,624	11,734	*97,330*
$25,000 – $49,999	28,550	51,655	14,926	22,146	*117,277*
$50,000 – $74,999	9,330	36,807	14,964	24,288	*85,389*
$75,000+	4,849	38,084	19,600	36,246	*98,779*
TOTAL HHS	*91,204*	*154,043*	*59,114*	*94,414*	***398,775***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	12.2%	6.9%	2.4%	2.9%	*24.4%*
$25,000 – $49,999	7.2%	13.0%	3.7%	5.6%	*29.4%*
$50,000 – $74,999	2.3%	9.2%	3.8%	6.1%	*21.4%*
$75,000+	1.2%	9.6%	4.9%	9.1%	*24.8%*
TOTAL HHS	*22.9%*	*38.6%*	*14.8%*	*23.7%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	46,947	26,630	9,321	11,364	*94,263*
$25,000 – $49,999	27,650	50,027	14,456	21,448	*113,581*
$50,000 – $74,999	9,036	35,647	14,492	23,523	*82,698*
$75,000+	4,696	36,884	18,982	35,104	*95,666*
TOTAL HHS	*88,330*	*149,188*	*57,251*	*91,439*	***386,208***

D: Sample Cell Targets (Region Sample Size = 1,200)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	146	83	29	35	*293*
$25,000 – $49,999	86	155	45	67	*353*
$50,000 – $74,999	28	111	45	73	*257*
$75,000+	15	115	59	109	*297*
TOTAL HHS	*274*	*464*	*178*	*284*	***1,200***

Appendix C-3
Northern Michigan Rural
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	48,689	25,571	7,424	8,065	*89,749*
$25,000 – $49,999	23,147	46,629	11,391	14,999	*96,166*
$50,000 – $74,999	7,126	29,098	10,041	13,813	*60,078*
$75,000+	3,198	26,803	11,229	18,289	*59,519*
TOTAL HHS	*82,160*	*128,101*	*40,085*	*55,166*	***305,512***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	15.9%	8.4%	2.4%	2.6%	*29.4%*
$25,000 – $49,999	7.6%	15.3%	3.7%	4.9%	*31.5%*
$50,000 – $74,999	2.3%	9.5%	3.3%	4.5%	*19.7%*
$75,000+	1.0%	8.8%	3.7%	6.0%	*19.5%*
TOTAL HHS	*26.9%*	*41.9%*	*13.1%*	*18.1%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	48,925	25,695	7,460	8,104	*90,185*
$25,000 – $49,999	23,259	46,855	11,446	15,072	*96,633*
$50,000 – $74,999	7,161	29,239	10,090	13,880	*60,370*
$75,000+	3,214	26,933	11,284	18,378	*59,808*
TOTAL HHS	*82,559*	*128,723*	*40,280*	*55,434*	***306,995***

D: Sample Cell Targets (Region Sample Size = 1,200)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	191	100	29	32	*353*
$25,000 – $49,999	91	183	45	59	*378*
$50,000 – $74,999	28	114	39	54	*236*
$75,000+	13	105	44	72	*234*
TOTAL HHS	*323*	*503*	*157*	*217*	***1,200***

Appendix C-4
Small Cities
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	28,640	12,095	4,829	5,015	*50,579*
$25,000 – $49,999	12,015	13,930	5,159	6,095	*37,199*
$50,000 – $74,999	3,765	9,125	4,325	5,315	*22,530*
$75,000+	1,379	8,429	4,878	7,474	*22,160*
TOTAL HHS	*45,799*	*43,579*	*19,191*	*23,899*	***132,468***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	21.6%	9.1%	3.6%	3.8%	*38.2%*
$25,000 – $49,999	9.1%	10.5%	3.9%	4.6%	*28.1%*
$50,000 – $74,999	2.8%	6.9%	3.3%	4.0%	*17.0%*
$75,000+	1.0%	6.4%	3.7%	5.6%	*16.7%*
TOTAL HHS	*34.6%*	*32.9%*	*14.5%*	*18.0%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	28,184	11,902	4,752	4,935	*49,773*
$25,000 – $49,999	11,824	13,708	5,077	5,998	*36,606*
$50,000 – $74,999	3,705	8,980	4,256	5,230	*22,171*
$75,000+	1,357	8,295	4,800	7,355	*21,807*
TOTAL HHS	*45,069*	*42,885*	*18,885*	*23,518*	***130,357***

D: Sample Cell Targets (Region Sample Size = 1,000)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	216	91	36	38	*382*
$25,000 – $49,999	91	105	39	46	*281*
$50,000 – $74,999	28	69	33	40	*170*
$75,000+	10	64	37	56	*167*
TOTAL HHS	*346*	*329*	*145*	*180*	***1,000***

Appendix C-5
GVMC
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	30,622	13,773	6,426	7,805	*58,626*
$25,000 – $49,999	22,198	25,518	9,116	12,789	*69,621*
$50,000 – $74,999	8,601	19,458	9,180	15,974	*53,213*
$75,000+	4,414	27,399	16,115	32,290	*80,218*
TOTAL HHS	*65,835*	*86,148*	*40,837*	*68,858*	***261,678***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	11.7%	5.3%	2.5%	3.0%	*22.4%*
$25,000 – $49,999	8.5%	9.8%	3.5%	4.9%	*26.6%*
$50,000 – $74,999	3.3%	7.4%	3.5%	6.1%	*20.3%*
$75,000+	1.7%	10.5%	6.2%	12.3%	*30.7%*
TOTAL HHS	*25.2%*	*32.9%*	*15.6%*	*26.3%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	30,819	13,861	6,467	7,855	*59,002*
$25,000 – $49,999	22,341	25,682	9,175	12,871	*70,069*
$50,000 – $74,999	8,657	19,583	9,239	16,077	*53,556*
$75,000+	4,442	27,575	16,219	32,498	*80,734*
TOTAL HHS	*66,259*	*86,701*	*41,100*	*69,301*	***263,361***

D: Sample Cell Targets (Region Sample Size = 1,000)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	117	53	25	30	*225*
$25,000 – $49,999	85	97	35	49	*266*
$50,000 – $74,999	33	74	35	61	*203*
$75,000+	17	105	61	123	*306*
TOTAL HHS	*252*	*329*	*156*	*263*	***1,000***

Appendix C-6
TCRPC
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	23,890	10,866	5,377	4,194	*44,327*
$25,000 – $49,999	16,190	16,412	5,642	6,693	*44,937*
$50,000 – $74,999	8,530	13,748	5,830	8,158	*36,266*
$75,000+	4,093	21,442	11,057	17,848	*54,440*
TOTAL HHS	*52,703*	*62,468*	*27,906*	*36,893*	***179,970***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	13.3%	6.0%	3.0%	2.3%	*24.6%*
$25,000 – $49,999	9.0%	9.1%	3.1%	3.7%	*25.0%*
$50,000 – $74,999	4.7%	7.6%	3.2%	4.5%	*20.2%*
$75,000+	2.3%	11.9%	6.1%	9.9%	*30.2%*
TOTAL HHS	*29.3%*	*34.7%*	*15.5%*	*20.5%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	24,370	11,085	5,485	4,278	*45,218*
$25,000 – $49,999	16,515	16,742	5,756	6,828	*45,841*
$50,000 – $74,999	8,702	14,025	5,947	8,322	*36,996*
$75,000+	4,175	21,873	11,279	18,207	*55,534*
TOTAL HHS	*53,762*	*63,725*	*28,467*	*37,635*	***183,589***

D: Sample Cell Targets (Region Sample Size = 800)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	106	48	24	19	*197*
$25,000 – $49,999	72	73	25	30	*200*
$50,000 – $74,999	38	61	26	37	*162*
$75,000+	18	95	49	79	*241*
TOTAL HHS	*234*	*277*	*124*	*165*	***800***

Appendix C-7
GCMPC
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	24,362	10,792	6,114	6,133	*47,401*
$25,000 – $49,999	16,331	17,956	5,807	7,358	*47,452*
$50,000 – $74,999	6,034	12,413	5,711	7,400	*31,558*
$75,000+	3,404	15,266	9,008	14,799	*42,477*
TOTAL HHS	*50,131*	*56,427*	*26,640*	*35,690*	***168,888***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	14.4%	6.4%	3.6%	3.6%	*28.1%*
$25,000 – $49,999	9.7%	10.6%	3.4%	4.4%	*28.1%*
$50,000 – $74,999	3.6%	7.3%	3.4%	4.4%	*18.7%*
$75,000+	2.0%	9.0%	5.3%	8.8%	*25.2%*
TOTAL HHS	*29.7%*	*33.4%*	*15.8%*	*21.1%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	24,407	10,812	6,126	6,144	*47,489*
$25,000 – $49,999	16,361	17,989	5,818	7,372	*47,540*
$50,000 – $74,999	6,045	12,436	5,722	7,414	*31,617*
$75,000+	3,410	15,294	9,025	14,827	*42,556*
TOTAL HHS	*50,223*	*56,531*	*26,691*	*35,757*	***169,202***

D: Sample Cell Targets (Region Sample Size = 800)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	115	51	29	29	224
$25,000 – $49,999	77	85	28	35	225
$50,000 – $74,999	29	59	27	35	150
$75,000+	16	72	43	70	201
TOTAL HHS	*237*	*267*	*127*	*169*	***800***

Appendix C-8
GLBR
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	22,362	10,628	4,759	4,073	*41,822*
$25,000 – $49,999	12,769	18,327	5,889	6,062	*43,047*
$50,000 – $74,999	4,993	12,156	5,229	7,059	*29,437*
$75,000+	2,459	14,697	8,180	14,701	*40,037*
TOTAL HHS	*42,583*	*55,808*	*24,057*	*31,895*	***154,343***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	14.5%	6.9%	3.1%	2.6%	*27.1%*
$25,000 – $49,999	8.3%	11.9%	3.8%	3.9%	*27.9%*
$50,000 – $74,999	3.2%	7.9%	3.4%	4.6%	*19.1%*
$75,000+	1.6%	9.5%	5.3%	9.5%	*25.9%*
TOTAL HHS	*27.6%*	*36.2%*	*15.6%*	*20.7%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	22,754	10,814	4,843	4,144	*42,555*
$25,000 – $49,999	12,993	18,649	5,992	6,168	*43,802*
$50,000 – $74,999	5,081	12,369	5,321	7,183	*29,954*
$75,000+	2,502	14,955	8,324	14,959	*40,740*
TOTAL HHS	*43,330*	*56,787*	*24,480*	*32,454*	***157,051***

D: Sample Cell Targets (Region Sample Size = 800)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	116	55	25	21	*217*
$25,000 – $49,999	66	95	31	31	*223*
$50,000 – $74,999	26	63	27	37	*153*
$75,000+	13	76	42	76	*207*
TOTAL HHS	*221*	*289*	*125*	*165*	***800***

Appendix C-9
WATS
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	16,135	7,115	2,765	2,515	*28,530*
$25,000 – $49,999	12,560	10,155	2,955	3,420	*29,090*
$50,000 – $74,999	6,370	9,185	3,510	4,310	*23,375*
$75,000+	5,670	19,545	11,160	16,785	*53,160*
TOTAL HHS	*40,735*	*46,000*	*20,390*	*27,030*	***134,155***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	12.0%	5.3%	2.1%	1.9%	*21.3%*
$25,000 – $49,999	9.4%	7.6%	2.2%	2.5%	*21.7%*
$50,000 – $74,999	4.7%	6.8%	2.6%	3.2%	*17.4%*
$75,000+	4.2%	14.6%	8.3%	12.5%	*39.6%*
TOTAL HHS	*30.4%*	*34.3%*	*15.2%*	*20.1%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	16,500	7,276	2,828	2,572	*29,176*
$25,000 – $49,999	12,844	10,385	3,022	3,497	*29,749*
$50,000 – $74,999	6,514	9,393	3,589	4,408	*23,904*
$75,000+	5,798	19,988	11,413	17,165	*54,364*
TOTAL HHS	*41,657*	*47,042*	*20,852*	*27,642*	***137,193***

D: Sample Cell Targets (Region Sample Size = 800)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	96	42	16	15	*170*
$25,000 – $49,999	75	61	18	20	*173*
$50,000 – $74,999	38	55	21	26	*139*
$75,000+	34	117	67	100	*317*
TOTAL HHS	*243*	*274*	*122*	*161*	***800***

Appendix C-10
KATS
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	16,410	7,377	3,326	3,247	*30,360*
$25,000 – $49,999	10,015	10,620	3,576	4,061	*28,272*
$50,000 – $74,999	3,457	7,945	3,524	4,658	*19,584*
$75,000+	2,368	11,572	6,233	10,784	*30,957*
TOTAL HHS	*32,250*	*37,514*	*16,659*	*22,750*	***109,173***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	15.0%	6.8%	3.0%	3.0%	*27.8%*
$25,000 – $49,999	9.2%	9.7%	3.3%	3.7%	*25.9%*
$50,000 – $74,999	3.2%	7.3%	3.2%	4.3%	*17.9%*
$75,000+	2.2%	10.6%	5.7%	9.9%	*28.4%*
TOTAL HHS	*29.5%*	*34.4%*	*15.3%*	*20.8%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	16,649	7,484	3,374	3,294	*30,801*
$25,000 – $49,999	10,161	10,774	3,628	4,120	*28,683*
$50,000 – $74,999	3,507	8,061	3,575	4,726	*19,869*
$75,000+	2,402	11,740	6,324	10,941	*31,407*
TOTAL HHS	*32,719*	*38,059*	*16,901*	*23,081*	***110,760***

D: Sample Cell Targets (Region Sample Size = 800)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	121	54	24	24	*223*
$25,000 – $49,999	73	78	26	30	*207*
$50,000 – $74,999	25	58	26	34	*143*
$75,000+	17	85	46	79	*227*
TOTAL HHS	*236*	*275*	*122*	*167*	***800***

Appendix C-11
WestPlan
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	12,755	6,187	2,504	2,702	*24,148*
$25,000 – $49,999	6,817	9,953	3,541	4,297	*24,608*
$50,000 – $74,999	2,267	6,800	3,205	4,665	*16,937*
$75,000+	1,017	8,204	3,872	7,652	*20,745*
TOTAL HHS	*22,856*	*31,144*	*13,122*	*19,316*	***86,438***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	14.8%	7.2%	2.9%	3.1%	*27.9%*
$25,000 – $49,999	7.9%	11.5%	4.1%	5.0%	*28.5%*
$50,000 – $74,999	2.6%	7.9%	3.7%	5.4%	*19.6%*
$75,000+	1.2%	9.5%	4.5%	8.9%	*24.0%*
TOTAL HHS	*26.4%*	*36.0%*	*15.2%*	*22.3%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	12,779	6,198	2,509	2,707	*24,193*
$25,000 – $49,999	6,830	9,972	3,548	4,305	*24,655*
$50,000 – $74,999	2,271	6,813	3,211	4,674	*16,969*
$75,000+	1,019	8,219	3,879	7,666	*20,783*
TOTAL HHS	*22,899*	*31,202*	*13,147*	*19,352*	***86,600***

D: Sample Cell Targets (Region Sample Size = 650)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	96	47	19	20	*182*
$25,000 – $49,999	51	75	27	32	*185*
$50,000 – $74,999	17	51	24	35	*127*
$75,000+	8	61	29	58	*156*
TOTAL HHS	*172*	*235*	*99*	*145*	***650***

Appendix C-12
JACTS
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	8,581	3,698	1,481	2,205	*15,965*
$25,000 – $49,999	4,693	7,003	2,265	2,528	*16,489*
$50,000 – $74,999	1,643	5,166	2,168	3,131	*12,108*
$75,000+	1,117	5,932	3,200	5,664	*15,913*
TOTAL HHS	*16,034*	*21,799*	*9,114*	*13,528*	***60,475***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	14.2%	6.1%	2.4%	3.6%	*26.4%*
$25,000 – $49,999	7.8%	11.6%	3.7%	4.2%	*27.3%*
$50,000 – $74,999	2.7%	8.5%	3.6%	5.2%	*20.0%*
$75,000+	1.8%	9.8%	5.3%	9.4%	*26.3%*
TOTAL HHS	*26.5%*	*36.0%*	*15.1%*	*22.4%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	8,623	3,716	1,488	2,216	*16,043*
$25,000 – $49,999	4,716	7,038	2,276	2,540	*16,570*
$50,000 – $74,999	1,651	5,191	2,179	3,146	*12,167*
$75,000+	1,122	5,961	3,216	5,692	*15,991*
TOTAL HHS	*16,112*	*21,906*	*9,159*	*13,594*	***60,771***

D: Sample Cell Targets (Region Sample Size = 650)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	92	40	16	24	*172*
$25,000 – $49,999	50	75	24	27	*176*
$50,000 – $74,999	18	56	23	34	*131*
$75,000+	12	64	34	61	*171*
TOTAL HHS	*172*	*235*	*97*	*146*	***650***

Appendix C-13a
TwinCATS
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	6,265	2,697	976	932	*10,870*
$25,000 – $49,999	2,931	3,251	1,184	1,222	*8,588*
$50,000 – $74,999	1,018	2,403	856	1,537	*5,814*
$75,000+	705	3,864	1,812	3,137	*9,518*
TOTAL HHS	*10,919*	*12,215*	*4,828*	*6,828*	***34,790***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	18.0%	7.8%	2.8%	2.7%	*31.2%*
$25,000 – $49,999	8.4%	9.3%	3.4%	3.5%	*24.7%*
$50,000 – $74,999	2.9%	6.9%	2.5%	4.4%	*16.7%*
$75,000+	2.0%	11.1%	5.2%	9.0%	*27.4%*
TOTAL HHS	*31.4%*	*35.1%*	*13.9%*	*19.6%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	6,119	2,635	953	910	*10,617*
$25,000 – $49,999	2,863	3,175	1,156	1,194	*8,388*
$50,000 – $74,999	994	2,347	837	1,501	*5,679*
$75,000+	689	3,774	1,770	3,064	*9,297*
TOTAL HHS	*10,665*	*11,931*	*4,716*	*6,669*	***33,981***

NOTE: Appendix C-13a is provided here for information only. The sample to be collected for the survey will be for the combined area as shown in Appendix C-13c.

Appendix C-13b
NATS
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	3,406	1,582	609	629	*6,226*
$25,000 – $49,999	1,622	2,575	761	1,381	*6,339*
$50,000 – $74,999	436	1,809	784	1,405	*4,434*
$75,000+	223	2,204	1,055	1,934	*5,416*
TOTAL HHS	*5,687*	*8,170*	*3,209*	*5,349*	***22,415***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	15.2%	7.1%	2.7%	2.8%	*27.8%*
$25,000 – $49,999	7.2%	11.5%	3.4%	6.2%	*28.3%*
$50,000 – $74,999	1.9%	8.1%	3.5%	6.3%	*19.8%*
$75,000+	1.0%	9.8%	4.7%	8.6%	*24.2%*
TOTAL HHS	*25.4%*	*36.4%*	*14.3%*	*23.9%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	3,547	1,647	634	655	*6,483*
$25,000 – $49,999	1,689	2,682	792	1,438	*6,601*
$50,000 – $74,999	454	1,884	816	1,463	*4,617*
$75,000+	232	2,295	1,099	2,014	*5,640*
TOTAL HHS	*5,922*	*8,508*	*3,341*	*5,570*	***23,341***

NOTE: APPENDIX C-13b is provided here for information only. The sample to be collected for the survey will be for the combined area as shown in APPENDIX C-13c.

Appendix C-13c
TwinCATS and NATS Combined
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	9,671	4,279	1,585	1,561	*17,096*
$25,000 – $49,999	4,553	5,826	1,945	2,603	*14,927*
$50,000 – $74,999	1,454	4,212	1,640	2,942	*10,248*
$75,000+	928	6,068	2,867	5,071	*14,934*
TOTAL HHS	*16,606*	*20,385*	*8,037*	*12,177*	***57,205***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	16.9%	7.5%	2.8%	2.7%	*29.9%*
$25,000 – $49,999	8.0%	10.2%	3.4%	4.6%	*26.1%*
$50,000 – $74,999	2.5%	7.4%	2.9%	5.1%	*17.9%*
$75,000+	1.6%	10.6%	5.0%	8.9%	*26.1%*
TOTAL HHS	*29.0%*	*35.6%*	*14.0%*	*21.3%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	9,666	4,282	1,587	1,565	*17,100*
$25,000 – $49,999	4,552	5,857	1,948	2,632	*14,989*
$50,000 – $74,999	1,448	4,231	1,653	2,964	*10,296*
$75,000+	921	6,069	2,869	5,078	*14,937*
TOTAL HHS	*16,587*	*20,439*	*8,057*	*12,239*	***57,322***

D: Sample Cell Targets (Region Sample Size = 800)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	135	60	22	22	*239*
$25,000 – $49,999	64	82	27	36	*209*
$50,000 – $74,999	20	59	23	41	*143*
$75,000+	13	85	40	71	*209*
TOTAL HHS	*232*	*286*	*112*	*170*	***800***

Appendix C-14
MACC
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	4,240	2,073	844	1,061	*8,218*
$25,000 – $49,999	3,400	4,660	1,474	2,936	*12,470*
$50,000 – $74,999	1,204	3,390	1,863	3,183	*9,640*
$75,000+	588	4,682	2,399	5,605	*13,274*
TOTAL HHS	*9,432*	*14,805*	*6,580*	*12,785*	***43,602***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	9.7%	4.8%	1.9%	2.4%	*18.8%*
$25,000 – $49,999	7.8%	10.7%	3.4%	6.7%	*28.6%*
$50,000 – $74,999	2.8%	7.8%	4.3%	7.3%	*22.1%*
$75,000+	1.3%	10.7%	5.5%	12.9%	*30.4%*
TOTAL HHS	*21.6%*	*34.0%*	*15.1%*	*29.3%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	4,255	2,080	847	1,065	*8,247*
$25,000 – $49,999	3,412	4,676	1,479	2,946	*12,513*
$50,000 – $74,999	1,208	3,402	1,869	3,194	*9,673*
$75,000+	590	4,698	2,407	5,624	*13,319*
TOTAL HHS	*9,465*	*14,856*	*6,602*	*12,829*	***43,752***

D: Sample Cell Targets (Region Sample Size = 650)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	63	31	12	16	*122*
$25,000 – $49,999	51	69	22	44	*186*
$50,000 – $74,999	18	51	28	47	*144*
$75,000+	9	70	35	84	*198*
TOTAL HHS	*141*	*221*	97	*191*	***650***

Appendix C-15
BCATS
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	6,062	2,361	1,276	1,056	*10,755*
$25,000 – $49,999	3,368	4,128	1,575	1,846	*10,917*
$50,000 – $74,999	1,219	2,798	1,301	1,718	*7,036*
$75,000+	728	3,326	1,703	3,009	*8,766*
TOTAL HHS	*11,377*	*12,613*	*5,855*	*7,629*	***37,474***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	16.2%	6.3%	3.4%	2.8%	*28.7%*
$25,000 – $49,999	9.0%	11.0%	4.2%	4.9%	*29.1%*
$50,000 – $74,999	3.3%	7.5%	3.5%	4.6%	*18.8%*
$75,000+	1.9%	8.9%	4.5%	8.0%	*23.4%*
TOTAL HHS	*30.4%*	*33.7%*	*15.6%*	*20.4%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	6,123	2,385	1,289	1,067	*10,864*
$25,000 – $49,999	3,402	4,169	1,591	1,864	*11,026*
$50,000 – $74,999	1,231	2,826	1,314	1,735	*7,106*
$75,000+	735	3,359	1,720	3,039	*8,853*
TOTAL HHS	*11,491*	*12,739*	*5,914*	*7,705*	***37,849***

D: Sample Cell Targets (Region Sample Size = 650)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	105	41	22	18	*186*
$25,000 – $49,999	58	72	27	32	*189*
$50,000 – $74,999	21	48	23	30	*122*
$75,000+	13	58	30	52	*153*
TOTAL HHS	*197*	*219*	*102*	*132*	***650***

Appendix C-16
TVC
(Household Size X Household Income)

A: 2006-2010 CTPP/ACS, Table A112208C – HH size by HH income in the past 12 months (2010$)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	4,140	1,628	609	323	*6,700*
$25,000 – $49,999	3,111	4,134	923	1,258	*9,426*
$50,000 – $74,999	1,360	3,305	968	1,657	*7,290*
$75,000+	693	4,445	1,737	3,287	*10,162*
TOTAL HHS	*9,304*	*13,512*	*4,237*	*6,525*	***33,578***

B: % Distribution of 2006-2010 CTPP/ACS Region HHs

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	12.3%	4.8%	1.8%	1.0%	*20.0%*
$25,000 – $49,999	9.3%	12.3%	2.7%	3.7%	*28.1%*
$50,000 – $74,999	4.1%	9.8%	2.9%	4.9%	*21.7%*
$75,000+	2.1%	13.2%	5.2%	9.8%	*30.3%*
TOTAL HHS	*27.7%*	*40.2%*	*12.6%*	*19.4%*	***100.0%***

C: Census 2010 HHs (based on CTPP/ACS 06-10 % distribution)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	4,184	1,645	616	326	*6,771*
$25,000 – $49,999	3,144	4,178	933	1,271	*9,526*
$50,000 – $74,999	1,374	3,340	978	1,675	*7,367*
$75,000+	700	4,492	1,755	3,322	*10,269*
TOTAL HHS	*9,402*	*13,655*	*4,282*	*6,594*	***33,933***

D: Sample Cell Targets (Region Sample Size = 650)

	1-person	2-person	3-person	4+ persons	TOT HH
< $25,000	80	32	12	6	*130*
$25,000 – $49,999	60	80	18	24	*182*
$50,000 – $74,999	26	64	19	32	*141*
$75,000+	13	86	34	64	*197*
TOTAL HHS	*179*	*262*	*83*	*126*	***650***

Appendix D

Substrata within Sample Areas

Appendix D
Substrata within Sample Areas

Table D-1. Substrata for SEMCOG minus WATS area

Sampling stratum	Description	HH income < $25K		HHs with 4+ persons		HHs with 0 autos		# Block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	54%	2%		1%		1%	43	1%
	H2: 4+ persons		12%	32%	32%		7%	640	19%
	H3: 0 autos		38%		16%	13%	61%	873	19%
	Total		51%		48%		68%	1556	38%
Low density	L: Other HHs		49%		52%		32%	2201	62%

Appendix D
Substrata within Sample Areas

Table D-2. Substrata for the Southern Michigan rural area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	32%	34%		19%	175	21%
	H2: 4+ persons		16%	28%	31%	173	22%
	Total		50%		50%	348	43%
Low density	L: Other HHs		50%		50%	486	57 %

Table D-3. Substrata for the Northern Michigan rural area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	41%	23%		12%	94	14%
	H2: 4+ persons		21%	21%	39%	160	26%
	Total		44%		51%	254	39%
Low density	L: Other HHs		56%		49%	450	61%

Table D-4. Substrata for the Small Cities area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	49%	34%		18%	67	22%
	H2: 4+ persons		14%	25%	32%	56	18%
	Total		48%		50%	123	40%
Low density	L: Other HHs		52%		50%	176	60%

Table D-5. Substrata for the GVMC area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	38%	37%		13%	80	16%
	H2: 4+ persons		13%	20%	37%	94	23%
	Total		49%		49%	174	39%
Low density	L: Other HHs		51%		51%	267	61%

Appendix D
Substrata within Sample Areas

Table D-6. Substrata for the TCRPC rural area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	41%	37%		14%	68	17%
	H2: 4+ persons		13%	26%	36%	85	23%
	Total		50%		50%	153	39%
Low density	L: Other HHs		50%		50%	218	61%

Table D-7. Substrata for the GCMPC area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	47%	38%		15%	88	18%
	H2: 4+ persons		12%	28%	34%	74	20%
	Total		50%		49%	162	38%
Low density	L: Other HHs		50%		51%	211	62%

Table D-8. Substrata for the GLBR area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	42%	36%		14%	66	17%
	H2: 4+ persons		14%	26%	33%	76	22%
	Total		50%		47%	142	39%
Low density	L: Other HHs		50%		53%	189	61%

Table D-9. Substrata for the WATS area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		HHs with 0 autos coverage	# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage			
High density	H1: Low income	40%	41%		12%	44%	46	17%
	H2: 4+ persons		9%	28%	38%	7%	45	20%
	Total		50%		50%	51%	91	37%
Low density	L: Other HHs		50%		50%	48%	160	63%

Appendix D
Substrata within Sample Areas

Table D-10. Substrata for the KATS area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	49%	36%		13%	40	17%
	H2: 4+ persons		13%	27%	36%	43	23%
	Total		49%		50%	83	40%
Low density	L: Other HHs		51%		50%	124	60%

Table D-11. Substrata for the WestPlan area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	40%	34%		11%	35	17%
	H2: 4+ persons		20%	29%	36%	40	22%
	Total		55%		47%	75	39%
Low density	L: Other HHs		45%		53%	100	61%

Table D-12. Substrata for the JACTS area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income		41%		17%	30	20%
	H2: 4+ persons		10%	27%	32%	26	20%
	Total	39%	51%		49%	56	40%
Low density	L: Other HHs		49%		51%	75	60%

Table D-13. Substrata for the TwinCATS+NATS area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	43%	37%		18%	25	17%
	H2: 4+ persons		13%	26%	30%	22	19%
	Total		50%		49%	47	36%
Low density	L: Other HHs		50%		51%	76	64%

Appendix D
Substrata within Sample Areas

Table D-14. Substrata for the MACC area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	33%	21%		8%	9	10%
	H2: 4+ persons		25%	36%	42%	20	30%
	Total		47%		50%	29	39%
Low density	L: Other HHs		53%		50%	49	61%

Table D-15. Substrata for the BCATS area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	44%	40%		15%	88	18%
	H2: 4+ persons		10%	28%	34%	74	20%
	Total		50%		49%	162	38%
Low density	L: Other HHs		50%		51%	211	62%

Table D-16. Substrata for the TVC area

Sampling stratum	Description	HH Income < $25K		HHs with 4+ persons		# block groups	Coverage of all HHs
		Minimum prevalence	Coverage	Minimum prevalence	Coverage		
High density	H1: Low income	35%	35%		10%	9	18%
	H2: 4+ persons		16%	25%	40%	14	23%
	Total		51%		50%	23	41%
Low density	L: Other HHs		48%		50%	31	59%

Appendix DDD

Final Weighting Plan

Westat®

Improving lives through research™

MDOT-MTC III Statewide Household Travel Survey

Final Weighting Plan

March 4, 2016

Prepared for:
Michigan Department of Transportation
Bureau of Transportation Planning
425 Ottawa Street
Lansing, MI 48909

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Table of Contents

Introduction and Background 1

1.1 Summary of 2015 Sample Design

The target population for the MTC III survey was all households in Michigan, where a household is defined as all people who occupy a housing unit. A housing unit may be a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied as separate living quarters. Separate living quarters are those in which the occupants live separately from any other individuals in the building and which have direct access from outside the building or through a common hall. The MTC III target population excluded individuals living in group quarters, where a group quarters is a place where people live or stay in group living arrangement that is owned or managed by an entity or organization providing housing and/or services for the residents. Group quarters include such places as college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, correctional facilities, and worker dormitories. The MTC III target population also excluded individuals staying in transient housing such as camp grounds, marinas, non-profit lodgings (YMCAs, youth hostels), and commercial hotels and motels.

The first level of sample stratification was the 16 geographical areas to be used to develop transportation planning models. These areas were defined by political boundaries, such as county boundaries or township lines. The less populous model areas were sampled at higher rates than the more populous areas to increase their sample sizes relative to the sample sizes that would have been produced by a simple proportional allocation.

Within each transportation model area, it was expected that some types of households would have lower response rates than other types of households. Consequently, within each primary stratum (i.e., each geographical sampling area), two substrata were created. One of the substrata, referred to as the **high-density substratum**, consisted of households that according to the 2008-2012 ACS were located in Census block groups that have a high density (relative to the other Census block groups) of characteristics associated with hard-to-reach households. The other sampling substratum, referred to as the **low-density substratum**, contained the households in all the other block groups in the sample area. For 14 of the primary strata the ACS characteristics used to define the high density substratum were household incomes less than $25,000, and households with 4 or more

persons. In the other two primary strata (WATS and SEMCOG Minus WATS), households with zero vehicles were also used to define the high density substratum.

Address-based sampling (ABS) was used to obtain a representative sample of household addresses for each of the 16 model areas and associated substrata. The sampling frame was a database of addresses created by Marketing Systems Group (MSG) from the U.S. Postal Service's (USPS) Computerized Delivery Sequence (CDS) file. The CDS file contains information on all delivery point addresses serviced by the USPS, with the exception of general delivery (mail held at a main post office for recipients to claim within 30 days). Within each model area, the high density stratum was oversampled relative to the low density stratum in order to increase the number of retrieval completes for hard-to-reach households. To permit telephone interviewers to contact nonresponding households, MSG used reverse telephone directories to match the sampled addresses to landline telephone numbers.

Two samples of addresses were selected for data collection during spring 2015 and fall 2015, respectively. The two samples were selected independently and the fall sample was de-duplicated so that a sampled household was asked to participate only once in the MTC III survey. (In the SEMCOG minus WATS area, additional SEMCOG-funded samples were also selected, which were de-duplicated from the MDOT-funded samples.) The same strata and substrata were used to select the spring and fall samples. In some of the model areas, the fall sample has a higher oversampling factor for the high-density substratum compared to the factor used in the spring, to increase the yield of hard-to-reach households.

The sample of addresses selected from each sample area and substratum was randomly partitioned into four subsamples, and the addresses in each subsample were assigned to the same travel day of a week (Monday through Thursday). After travel day assignment, 40 percent of the sampled addresses were randomly flagged for invitation to participate in the GPS subsample.

Table 1-1 contains the names of the model areas associated with the primary sampling strata, shows the number of addresses in the MSG sampling frame used to select the fall sample, and specifies the sizes of the fielded spring and fall samples.

1.2 Data collection results

Table 1-1 also shows the overall response rates by individual substrata for the MDOT-funded samples.

Table 1-1. Sampling strata and substrata for MDOT-funded samples

Sample area	Housing units (occupied)[2]	Sub-stratum	Addresses in MSG frame[3]	Fielded sample sizes			Overall response Rate
				Spring	Fall	Both	
Statewide model							
Southeast Michigan Council of Governments (SEMCOG) minus Washtenaw County (WATS)[1]	1,707,565	Low	1,120,856	10,396	11,836	22,232	4.47%
		High	680,266	7,492	8,401	15,893	5.40%
Southern Michigan Rural	386,208	Low	233,159	5,234	4,664	9,898	6.66%
		High	170,737	5,486	4,900	10,386	6.54%
Northern Michigan Rural (NMR)	306,995	Low	193,842	5,115	3,833	8,948	7.67%
		High	123,895	4,658	3,527	8,185	7.01%
Small Cities	130,357	Low	86,941	4,433	4,267	8,700	6.49%
		High	56,297	3,837	3,684	7,521	7.31%
Urban model areas							
Grand Valley Metropolitan Council (GVMC)	263,361	Low	171,593	5,239	2,267	7,506	7.62%
		High	113,979	3,822	1,653	5,475	8.65%
Tri-County Regional Planning Commission (TCRPC)	183,589	Low	117,432	3,765	1,469	5,234	7.67%
		High	78,108	3,605	1,412	5,017	8.65%
Genesee County Metropolitan Planning Commission (GCMPC)	169,202	Low	108,562	3,971	4,587	8,558	5.59%
		High	61,203	3,042	5,961	9,003	5.90%
Great Lakes Bay Region (GLBR)	157,051	Low	98,811	4,204	2,668	6,872	7.00%
		High	63,947	3,162	2,011	5,173	7.77%
Washtenaw Area Transportation Study (WATS)	137,193	Low	91,208	3,403	2,097	5,500	6.74%
		High	55,959	3,744	3,346	7,090	8.36%
Kalamazoo Area Transportation Study (KATS)	110,760	Low	71,732	3,717	2,666	6,383	6.39%
		High	48,964	3,356	2,400	5,756	7.97%
West Michigan Metropolitan Transportation Planning Program (WestPlan)	86,600	Low	58,131	2,914	2,894	5,808	5.54%
		High	38,121	2,806	2,751	5,557	7.58%
Jackson Area Comprehensive Transportation Study (JACTS)	60,771	Low	38,216	2,752	4,635	7,387	5.51%
		High	27,710	2,926	4,966	7,892	7.00%
Twin Cities Area Transportation Study (TwinCATS) and Niles/Buchanan/Cass Area Transportation Study (NATS)[2]	57,322	Low	38,949	3,928	3,955	7,883	5.23%
		High	23,294	3,115	3,190	6,305	7.40%
Macatawa Area Coordinating Council (MACC)	43,752	Low	30,032	3,425	3,708	7,133	6.51%
		High	18,574	2,232	2,438	4,670	8.47%
Battle Creek Area Transportation Study (BCATS)	37,849	Low	24,487	3,196	4,685	7,881	5.32%
		High	15,704	2,746	3,892	6,638	6.19%
Traverse City (TVC)	33,933	Low	22,159	2,549	808	3,357	8.46%
		High	15,771	3,390	1,030	4,420	8.61%
TOTAL	**3,872,508**		4,098,639	127,660	116,601	244,261	

[1] The SEMCOG minus WATS sample size is based upon the Statewide Model needs.

[2] Source: 2010 Decennial Census

[3] MSG sampling frames used for selection of fall sample

Weighting Concepts for Calculation of Household Weights 2

2.1 Types of Weights

When different strata have different sampling rates or different response rates, it is necessary to use weights[1] when analyzing the collected data. If weighting is not conducted, then areas and sub-areas with higher sampling rates or higher response rates will be over-represented relative to areas and sub-areas with lower response rates. Because survey data can be reported at different levels—for example, household level or person level—weights are often labeled with their associated reporting unit. For example, *household weights* are used to compute estimates from data reported at the household level, whereas *person weights* are used to compute estimates from data reported at the person level.

Weights used in the calculation of point estimates (such as means, totals, and regression estimates) are also referred to as *full-sample weights* because they expand data obtained from all of the sample's responding households to the universe of all households in the sampling area. In a data set containing survey data and associated full-sample household weights, there is one full-sample household weight for each responding household. Full-sample weights are so named because there is no subsampling performed prior to calculating full-sample weights. However, separate sets of full-sample weights may be calculated for different sub-populations, such as for the 16 MDOT model areas. Given a set of full-sample weights for a particular sub-population, however, it is not necessary to calculate new weights for smaller sub-populations within the particular sub-population. Instead, indicator variables equal to 1 for the respondents belonging to the smaller sub-populations and equal to 0 for the respondents that do not belong to the smaller sub-populations can be incorporated into estimates of means, totals, and regression coefficients to estimate modeling parameters for the smaller sub-populations.

Replicate weights are used to compute measures of precision, such as standard errors and confidence intervals. The calculation of replicate weights involves computing weights for random subsamples

[1] Also referred to as *household expansion factors*.

of the full sample. In a data set of survey data and associated weights, there is a set of replicate weights for each respondent. Various methods are used to randomly subsample the full sample to create a set of replicate weights. We plan to use the paired jackknife replication method, which for each calculated replicate weight selects a subsample that is one half the size of the full sample. We plan to compute 100 replicate weights for each respondent.

For the MDOT-funded MTC III samples, Westat plans to deliver separate files of household weights and person weights for the retrieval completes. The file of household weights will contain a full-sample household weight and an associated set of replicate household weights for each household in the MDOT-funded sample that completed a retrieval interview. The file of person weights will contain a full-sample person weight and an associated set of replicate person weights for each eligible person in a household in the MDOT-funded sample that completed a retrieval interview.

Weights are usually developed in a series of steps to compensate for unequal selection probabilities, nonresponse, non-coverage, and sampling fluctuations from known population totals. The steps in the weighting process will be the following:

1. Construction of household base weights (the reciprocal of the probability of selection of each sampled address);

2. Adjustment for non-response at the household level, yielding nonresponse-adjusted weights;

3. Further adjustment of the household weights to achieve consistency with characteristics for the full population of households (achieved by *poststratifying* or *raking* the non-response adjusted weights to independent household level totals, referred to as *control totals*).

4. Calculation of initial person weights, which will be the nonresponse-adjusted household weight applied to each eligible person in a responding household; and

5. Further adjustment of the initial person weights to achieve consistency with characteristics for the universe of eligible persons (achieved by poststratifying or raking the initial person-level weights to independent person-level control totals).

We plan to include a summary of the final weights in the weighting section of the final project report. The summary will include tables containing the values of the control totals; descriptive statistics of the final weights, such as sums of weights, averages of the weights, and maximum values of the weights; and comparisons of weighted and unweighted statistics for key survey statistics.

2.2 Calculation of Base Weights

The household base weight reflects the probability of selection for a sampled household and is calculated simply as the reciprocal of its probability of selection. That two samples were selected—a spring sample and a fall sample—complicates the calculation of the base weights because MSG updated their address frame several times between the selection of the spring and fall samples. If only one sample had been selected, the formula for the base weight would be the following:

$$Base\ weight = N_h / n_h ,$$

where N_h is the number of addresses in MSG's sampling frame for stratum h, and n_h is the number of addresses sampled from stratum h. However, because two samples were selected, the denominator of the base weight becomes the sum of the unduplicated number of addresses selected in either the spring or fall. For the numerator, some surveys would use the stratum's average of the spring and fall frame sizes. ***We propose, instead, to use the fall sample's frame size for the numerator of the base weight because the date of the address frame used to select the fall sample is closer to the middle of 2015 than that of the address frame used to select the spring sample and should thus better represent the universe of addresses in 2015***. Base weights will be calculated for all sampled addresses, not just for respondents.

Table 2-1 contains the calculated base weights for each sample area and sub-stratum for the MDOT-funded samples.

2.3 Adjustments for Household-Level Non-Response

At the end of data collection, each sampled address will be assigned a disposition code, which for weighting purposes can be assigned to one of the following four categories:

- *Undeliverable address;*
- *Eligible respondent*, i.e. a retrieval complete*;*

- *Other-ineligible case (i.e., other than being an undeliverable address),* e.g., a business address instead of a residential address; and
- *Non-respondent*, e.g., (1) a recruitment refusal, (2) a recruitment complete that fails to complete a retrieval interview, (3) a non-respondent to the recruitment interview, with no evidence that the address was undeliverable.

Table 2-1. Base weights for MDOT-funded samples

Sample area	Base weight by sub-stratum	
	Low density	High density
Statewide models		
SEMCOG minus WATS	50.416	42.803
Southern Michigan Rural	23.556	16.439
Northern Michigan Rural	21.663	15.137
Small Cities	9.993	7.485
Urban model areas		
GVMC	22.861	20.818
TCRPC	22.436	15.569
GCMPC	12.685	6.798
GLBR	14.379	12.362
WATS	16.583	7.893
KATS	11.238	8.507
WestPlan	10.009	6.860
JACTS	5.173	3.511
TwinCATS + NATS	4.941	3.695
MACC	4.210	3.977
BCATS	3.107	2.366
TVC	6.601	3.568

The base weights of the *eligible respondents* will be multiplied by a nonresponse adjustment factor to compensate for the non-completion of recruitment interviews by the *non-respondents*. The base weights for the *other-ineligible cases* will also be adjusted to reflect that some *non-respondents* may be ineligible. The base weights for the *undeliverable addresses* will not be adjusted, and the adjusted base weights for the *non-respondents* will be set to zero.

The calculation of the nonresponse adjustment factors requires that each sampled address that is not an undeliverable address is assigned to a set of sampled addresses, called a *nonresponse adjustment cell*. Sampled addresses assigned to the same nonresponse adjustment cell should have similar response rates. Both respondents and non-respondents are assigned to nonresponse adjustment cells, so the assignment process must be based on information available for both respondents and non-

respondents. ***We plan to use the 32 sampling substrata to define the nonresponse adjustment cells and will investigate also using wave (i.e. spring or fall), match status (i.e., whether or not an address was matched to a landline telephone) and, as suggested by Norman and Sigman[2], the type-of delivery point (i.e., city delivery, rural delivery, P.O. boxes, and highway contact) to define nonresponse adjustment cells***. So that the adjusted base weights for MDOT-funded retrieval completes in the SEMCOG Minus WATS area can be used to calculate composite weights for MDOT-plus-SEMCOG-funded retrieval completes, the county containing the sampled address will also be used to define nonresponse adjustment cells for the MDOT-funded retrieval completes in the SEMCOG Minus WATS area.

After the nonresponse adjustment cells are created, the following formula will be used to compute nonresponse adjustment factors for each cell:

$$f_c^{(1)} = \frac{\sum_{i \in ER_c} w_i + \sum_{i \in OI_c} w_i + \sum_{i \in NR_c} w_i}{\sum_{i \in ER_c} w_i + \sum_{i \in OL_c} w_i},$$

where $\sum_{i \in ER_c} w_i$ is the sum of base weights for *eligible respondents* in weighting cell *c,* $\sum_{i \in OI_c} w_i$ is the sum of base weights for *other-ineligible cases* in weighting cell *c,* and $\sum_{i \in NR_c} w_i$ is the sum of base weights for *non-respondents* in weighting class c .

2.4 Poststratifying or Raking at the Household Level

Poststratification and raking are procedures used to improve the reliability of survey estimates and, to some extent, to correct for the bias due to under-coverage and/or non-response. These procedures further adjust the nonresponse-adjusted weights so that within individual demographic categories the sum of the further adjusted nonresponse-adjusted weights for retrieval completes equals a corresponding independent control total for the category. To illustrate the difference between poststratification and raking, assume that independent control totals are available for two

[2] Norman, G. and Sigman, R. (2007). Using addresses as sampling units in the 2007 Health Information National Trends Survey, Proceedings of the Survey Research Section, American Statistical Association, Alexandria, VA.

demographic variables, which we will call the row dimension and the column dimension. For poststratification, it is necessary that the independent control totals be known for the cross-classification of the two demographic variables:

T_{ij} = control total for category *i* of the row dimension and category *j* of the column dimension.

For raking, on the other hand, it is necessary only that the independent control totals be known for each category of each variable, not for the cross-classification of the two variables:

T_{i+} = control total for category *i* of the row dimension, and

T_{+j} = control total for category *j* of the column dimension.

Continuing the example of having independent control totals for two demographic variables, poststratification further adjusts the nonresponse-adjusted weights for the retrieval completes by multiplying them by the following adjustment factor:

$f_{ij}^{(2)}$ = poststratification factor for weights associated with category *i* of the row dimension and category *j* of the column dimension

$= T_{ij}/S_{ij}$,

where S_{ij} is the sum of the nonresponse-adjusted weights of the retrieval completes associated with category *i* of the row demographic variable and category *j* of the column demographic variable. Raking, on the other hand, further adjusts the nonresponse-adjusted weights for the retrieval completes by multiplying them by the product of the following two adjustment factors:

$f_{i+}^{(2)}$ = raking factor for weights associated with category *i* of the row dimension, and

$f_{+j}^{(3)}$ = raking factor for weights associated with category *j* of the column dimension.

Raking is repeated poststratification relative to the individual dimensions. The first step is to update the nonresponse-adjusted weights for the retrieval completes by poststratifying them relative to the row dimension. The resulting updated weights are then further updated by poststratifying them relative to the column dimension. This process is then repeated on successively updated versions of the weights until, for all the categories of each dimension, the sum of the updated weights for the retrieval completes equals the independent control total for the category (within a small tolerance specified for convergence.)

Though the above discussion is for two demographic variables, raking and poststratification can be performed for more than two demographic variables. As discussed above, the choice between using poststratification or raking to benchmark weights to external control totals depends on whether the control totals are available for multi-dimensional categories resulting from a cross-classification of the demographic variables or only for the single-dimensional categories of each demographic variable. The MTC III RFP states that the household weights for the MDOT-funded samples are to

be benchmarked to the 2010 Census within three-dimensional cells defined by the 16 model areas, four levels of household income (less than $25,000, $25,000 to $49,999, $50,000 to $74,999 and $75,000 or more), and four levels of the number of persons in the household (1, 2, 3, and 4 or more). Section 3 discusses data sources for the external totals and various options for using these data sources to benchmark the MTC III household weights.

2.5 Imputation

The demographic variables used for poststratification or raking require valid values for every retrieval complete. Of the three variables that will be used to benchmark the MTC III weights—model area, household income level, and number of persons in the household—only household income may be missing for some retrieval completes. Before performing poststratification or raking, we plan to use *hot-deck* imputation to replace missing values of household-income level with valid values that can be used in poststratification or raking (i.e., less than $25,000, $25,000 to $49,999, $50,000 to $74,999 and $75,000 or more). The general approach of hot-deck imputation is to match records with missing survey data for a particular item to records with reported data for that item, referred to as *donor records*, using a set of matching variables; and then to assign each record with missing data a valid value for the item from a randomly selected matched donor record. The individual categories of the matching variables can be defined to be *soft boundaries* or *hard boundaries*. If a category is a soft boundary, then when matching a record that has missing data to the donor records if no donors can be found the matching procedure will "reach across" to adjacent categories of the matching variable to find donors. For example, if the number of vehicles is a matching variable, and a record with five vehicles cannot be matched to some donor record then the matching procedure will attempt to find donors with four or six vehicles; if there are none, then three or seven vehicles, etc. If, on the other hand, a category is a hard boundary, then the matching procedure is not allowed to reach across to adjacent categories. For imputing missing values of household-income level, we plan to investigate using number of workers and number of vehicles as soft-boundary matching variables and substratum, match status, type of delivery point, and ZIP code as hard-boundary matching variables. For the SEMCOG minus WATS area, county will also be a hard-boundary matching variable.

3 Data Sources for Control Totals

3.1 Benchmarking of Household Weights

The MTC III RFP states that the household weights for the MDOT-funded samples are to be benchmarked to the 2010 Census within three-dimensional cells defined by the 16 model areas, four levels of household income (less than $25,000, $25,000 to $49,999, $50,000 to $74,999 and $75,000 or more), and four levels of the number of persons in the household (1, 2, 3, and 4 or more). Tables in the appendices of the RFP, and also in Appendices C-1 through C-16 of the Westat-prepared MTC III sampling plan, contain a set of associated control totals. Each appendix contains four tables, labeled A, B, C, and D. Table A in each appendix contains for an individual model area estimated household counts for the cross tabulation of the four levels of household income with the four levels of household size. The Table A household counts are estimated from the Census Bureau's Census Transportation Planning Products (CTPP) special tabulation of 2006-2010 five-year ACS data. Table B in each appendix contains the distribution percentages corresponding to the Table A household counts, which are applied to the model area's number of households according to the 2010 Census to yield Table C, the model area's number of households according to the 2010 Census disaggregated by the cross tabulation of the four levels of household income with the four levels of household size. The remainder of this section discusses whether the Table C counts in the RFP and sampling plan should be used as the control totals to benchmark the MTC III weights, or if these counts should be updated using more recent ACS data.

Appendix A of this report compares for each model area one-way distribution percentages for household income levels and household sizes computed from 2006-2010 CTPP/ACS data with corresponding estimates computed from 2009-2013 ACS data. Differences between results computed with 2006-2010 data versus those computed with 2009-2013 data are highlighted if the difference exceeds 1 percentage point in absolute value. Out of 16 x (4+4) = 128 such differences, only 26 are greater than 1 percentage point in absolute value. For example, for the WATS area the largest absolute differences in one-way distribution percentages were for the household income level of $50,000 to $74,999 (an absolute change of 1.1 percentage point) and for one-person households (also an absolute difference of 1.1 percentage points). To determine whether five-year ACS estimates are sufficiently precise so that these differences indicate that purposes of computing control totals the distribution percentages for the WATS area should be based on 2009-2013 ACS

data instead of 2006-2010 ACS data, one can examine the ACS's margins of error for 2006-2010 estimates, for 2009-2013 estimates, and for the differences between these estimates. This is fairly easy to do for the WATS area because it is a single county, and one can use the Census Bureau's American Fact Finder web application to obtain margins of error for associated ACS estimates. Obtaining margins of errors for ACS estimates for other model areas, however, is much more difficult because margins of error for ACS estimates cannot be obtained from American Fact Finder for groups of counties or groups of Census block groups.

Table 3.1 contains estimates and changes in estimates, plus the associated margins of errors for 90% confidence intervals, for the distribution percentages for household-income levels in the WATS area, computed with 2006-2010 ACS data and with 2009-2013 ACS data. Table 3.1 indicates that for the income level of $50,000 to $74,000 the margin of error for a 90% confidence interval about the estimated 1.1 percentage point difference between the 2006-2010 period and the 2009-2013 period is 1 percentage point. In other words, the 90% confidence interval for the difference is from 0.1 percentage points to 2.1 percentage points. (The 95% confidence interval is slightly wider and includes zero.)

Table 3-1. Estimates and changes in estimates for one-way distribution percentages of WATS household-income levels. Associated margins of error (MOEs) for 90% confidence intervals are shown in parentheses.

Category	Estimated percentage (and MOE)		
	2006-2010 ACS	2009-2013 ACS	Change
Less than $25,000	21.3% (+/-0.9%)	21.3% (+/-0.7%)	0.0% (+/-1.2%)
$25,000 - $49,999	21.7% (+/-0.7%)	21.9% (+/-0.6%)	0.2% (+/-1.0%)
$50,000 - $74,999	17.4% (+/-0.6%)	16.4% (+/-0.7%)	-1.1% (+/-1.0%)
$75,000 or more	39.6% (+/-0.9%)	40.4% (+/-0.9%)	0.8% (+/-1.2%)

Seven of the MTC III urban model areas contain fewer households than the WATS model area. These smaller model areas will have smaller ACS sample sizes and thus the margins of error for associated ACS estimates will be larger. The smallest MTC III model area is Traverse City (TVC). Because it contains parts of two different counties, one cannot use American Fact Finder to obtain margins of error for ACS estimates for the TVC area. However, Midland County, Michigan, has approximately the same number of households as the TVC area (i.e., approximately 35,000 households). Table 3-2 compares for the TVC area and for Midland County estimates and changes in estimates, along with margins of error for Midland County, the distribution percentages of household-income levels, computed with 2006-2010 ACS data and with 2009-2013 ACS data. The

two largest absolute changes in the estimated distribution percentages of household-income levels for the TVC area were 1.8 and 2.1 percentage points, but as indicated by the corresponding margins of error for Midland County a major portion of these changes could be the sampling variability of ACS estimates for areas containing approximately 35,000 households.

Table 3-2. Estimates and changes in estimates for one-way distribution percentages of household-income levels in the TVC area and in Midland County. Associated margins of error (MOEs) for 90% confidence intervals are shown in parentheses.

	Estimated percentage (and MOE)					
	TVC area			Midland County		
Category	2006-2010 ACS	2009-2013 ACS	Change	2006-2010 ACS	2009-2013 ACS	Change
Less than $25,000	20.0%	20.9%	1.0%	23.8% (+/-1.4%)	22.6% (+/-1.5%)	-1.2% (+/-2.1%)
$25,000 - $49,999	28.1%	25.9%	-2.1%	25.3% (+/-1.3%)	24.8% (+/-1.4%)	-0.4% (+/-1.9%)
$50,000 - $74,999	21.7%	21.1%	-0.6%	18.1% (+/-1.3%)	19.2% (+/-1.2%)	1.1% (+/-1.8%)
$75,000 or more	30.3%	32.1%	1.8%	32.8% (+/-1.7%)	33.4% (+/-1.5%)	0.6% (+/-2.3%)

The examination of the margins of error for ACS estimates appears to suggest that the accuracy gains from using more recent ACS data to determine distribution percentages may be small because of the amount of sampling variability present in ACS estimates. It should be kept in mind, however, that five-year ACS estimates involving income are in terms of dollars for the last year of the five year period. Thus, estimated percentages for household income levels computed from 2006-2010 ACS data are in terms of 2010 dollars, whereas estimates computed from 2010-2014 ACS data are based on 2014 dollars. Both poststratification and raking require that the categories for the benchmarking variables be determined from the survey questions (or the sampling frame data) in the same way they are determined from the external data used compute the control totals. ***Consequently, we recommend that the most recent available ACS data be used in determining the MTC III control totals—that is, the control totals should be the result of using the most recent ACS data to distribute the total number of households in a model area according to the 2010 Census to within-model-area cells defined by household income and the number of persons in a household.***

According to the Census Bureau's ACS release schedule, the summary data for 2010-2014 ACS became available on December 3, 2015, and the corresponding public use file was available on January 21, 2016. However, the public use-file for 2010-2014 ACS data cannot be used to compute

MTC III control totals because the most-detailed geographical identifier on the public use file is for Public Use Microdata Areas (PUMAs), which for disclosure-avoidance reasons must contain at least 100,000 people and do not line up with the MTC III model areas.

By using the shape files for the model areas, however, one can determine a set of Census block groups associated with each model area. These can then be applied to the ACS summary data to compute estimates for each model area. The ACS summary data do not contain cross tabulations of household income levels with household size for individual block group, but one can obtain one-way distributions of household income levels and household size at the block-group level, which can then be aggregated to model areas.

One approach to using 2010-2014 ACS summary data to benchmark MTC III weights is to rake the weights using two dimensions: (1) model area crossed with household income levels and (2) model area crossed with household size. The control totals for each model area would be the area's number of households according to the 2010 Census multiplied by the one-way distribution percentages for household income levels and for household size computed from the 2010-2014 ACS summary data. This approach will be referred to as the *weight-raking* approach.

A second, more time consuming approach is to rake the RFP's Table C household counts, which are cross-classification counts by household income level and household size in each model area, to the two dimensions of control totals used in the weight raking approach. The updated cross-classified counts would then be used to poststratify the weights in each model area. This approach will be referred to as the *count-raking* approach.

We believe that the count-raking approach is the preferred approach because it obtains information about the relationships between household income levels (in 2010 dollars) and household size from the Census Bureau's CTPP special tabulation of 2006-2010 ACS data and then adjusts these relationships using more recent ACS data. The weight raking approach, on the other hand, obtains initial information about the relationships between household income levels (in 2015 dollars) and household size from the survey responses of the retrieval completes, but this information is affected by the dependence of response propensity on household income levels and family size. ***We plan to implement the count raking approach, but if unforeseen and unsolvable technical difficulties occur we will implement the weight raking approach as a fall**back.*

3.2 Benchmarking of Person Weights

In each model area, poststratification will be used to benchmark the person weights. The poststratified weights for all the respondents in a model area will sum to the number of persons in the area according to the 2010 Census. Similarly, the poststratified weights for all respondents age 15 or younger in each model area will sum to the number of persons age 15 or younger in the area according to the 2010 Census. The person weights for respondents age 16 or older will be poststratified to counts of persons from the 2010 Census that have been redistributed to poststratification cells defined by Sex (male or female), Age (16-19, 20-54, 55-64, or 65+), and Employment Status (civilian employed, civilian unemployed or not in the labor force, or in the armed forces), based on estimates in Table S2301 of the 2010-2014 five-year ACS summary data.

Appendix A
Comparison of 2006-2010 CTPP with 2009-2013 ACS

(Shaded Distribution-Change Cells are at Least 1% in Absolute Value)

A. SEMCOG Minus WATS

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	398,925	23.6%	415,402	24.7%	1.1%
$25,000 – $49,999	409,475	24.2%	398,234	23.7%	-0.5%
$50,000 – $74,999	306,600	18.1%	293,900	17.5%	-0.6%
$75,000+	577,485	34.1%	573,373	34.1%	0.0%
1-person	497,765	29.4%	500,478	29.8%	0.4%
2-person	536,800	31.7%	536,322	31.9%	0.2%
3-person	262,155	15.5%	260,318	15.5%	0.0%
4+person	395,765	23.4%	383,791	22.8%	-0.6%
TOTAL HHS	1,692,485	100.0%	1,680,909	100.0%	

B. Southern Michigan Rural

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	97,330	24.4%	94,084	24.7%	0.2%
$25,000 – $49,999	117,277	29.4%	110,253	28.9%	-0.5%
$50,000 – $74,999	85,389	21.4%	79,837	20.9%	-0.5%
$75,000+	98,779	24.8%	97,450	25.5%	0.8%
1-person	91,204	22.9%	91,285	23.9%	1.0%
2-person	154,043	38.6%	148,322	38.9%	0.2%
3-person	59,114	14.8%	56,544	14.8%	0.0%
4+person	94,414	23.7%	85,473	22.4%	-1.3%
TOTAL HHS	398,775	100.0%	381,624	100.0%	

C. Northern Michigan Rural

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	89,749	29.4%	88,802	29.2%	-0.1%
$25,000 – $49,999	96,166	31.5%	93,529	30.8%	-0.7%
$50,000 – $74,999	60,078	19.7%	59,056	19.4%	-0.2%
$75,000+	59,519	19.5%	62,428	20.5%	1.1%
1-person	82,160	26.9%	85,000	28.0%	1.1%
2-person	128,101	41.9%	128,456	42.3%	0.4%
3-person	40,085	13.1%	38,427	12.6%	-0.5%
4+person	55,166	18.1%	51,932	17.1%	-1.0%
TOTAL HHS	305,512	100.0%	303,815	100.0%	

Appendix A (continued)
Comparison of 2006-2010 CTPP with 2009-2013 ACS
(Shaded Distribution-Change Cells are at Least 1% in Absolute Value)

D. Small Cities

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	50,579	38.2%	47,687	37.5%	-0.7%
$25,000 – $49,999	37,199	28.1%	35,660	28.0%	-0.1%
$50,000 – $74,999	22,530	17.0%	21,637	17.0%	0.0%
$75,000+	22,160	16.7%	22,351	17.6%	0.8%
1-person	45,799	34.6%	43,906	34.5%	-0.1%
2-person	43,579	32.9%	42,451	33.3%	0.4%
3-person	19,191	14.5%	18,096	14.2%	-0.3%
4+person	23,899	18.0%	22,882	18.0%	-0.1%
TOTAL HHS	132,468	100.0%	127,335	100.0%	

E. GVMC

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	58,626	22.4%	57,402	21.6%	-0.8%
$25,000 – $49,999	69,621	26.6%	68,449	25.8%	-0.9%
$50,000 – $74,999	53,213	20.3%	53,864	20.3%	-0.1%
$75,000+	80,218	30.7%	86,090	32.4%	1.7%
1-person	65,835	25.2%	66,309	24.9%	-0.2%
2-person	86,148	32.9%	89,111	33.5%	0.6%
3-person	40,837	15.6%	42,094	15.8%	0.2%
4+person	68,858	26.3%	68,291	25.7%	-0.6%
TOTAL HHS	261,678	100.0%	265,805	100.0%	

F. TCRPC

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	44,327	24.6%	44,771	24.8%	0.1%
$25,000 – $49,999	44,937	25.0%	45,761	25.3%	0.4%
$50,000 – $74,999	36,266	20.2%	34,327	19.0%	-1.2%
$75,000+	54,440	30.2%	55,860	30.9%	0.7%
1-person	52,703	29.3%	54,434	30.1%	0.8%
2-person	62,468	34.7%	63,759	35.3%	0.6%
3-person	27,906	15.5%	26,323	14.6%	-0.9%
4+person	36,893	20.5%	36,203	20.0%	-0.5%
TOTAL HHS	179,970	100.0%	180,719	100.0%	

Appendix A (continued)
Comparison of 2006-2010 CTPP with 2009-2013 ACS
(Shaded Distribution-Change Cells are at Least 1% in Absolute Value)

G. GCMPC

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	47,401	28.07%	48,861	29.5%	1.4%
$25,000 – $49,999	47,452	28.10%	46,526	28.1%	0.0%
$50,000 – $74,999	31,558	18.7%	29,686	17.9%	-0.8%
$75,000+	42,477	25.2%	40,596	24.5%	-0.6%
1-person	50,131	29.7%	49,862	30.1%	0.4%
2-person	56,427	33.4%	56,264	34.0%	0.6%
3-person	26,640	15.8%	25,908	15.6%	-0.1%
4+person	35,690	21.1%	33,635	20.3%	-0.8%
TOTAL HHS	168,888	100.0%	165,669	100.0%	

H. GLBR

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	41,822	27.1%	41,884	27.0%	-0.1%
$25,000 – $49,999	43,047	27.9%	42,738	27.5%	-0.4%
$50,000 – $74,999	29,437	19.1%	29,367	18.9%	-0.2%
$75,000+	40,037	25.9%	41,267	26.6%	0.6%
1-person	42,583	27.6%	45,370	29.2%	1.6%
2-person	55,808	36.2%	55,644	35.8%	-0.3%
3-person	24,057	15.6%	22,859	14.7%	-0.9%
4+person	31,895	20.7%	31,383	20.2%	-0.5%
TOTAL HHS	154,343	100.0%	155,256	100.0%	

I. WATS

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	28,530	21.3%	28,904	21.3%	0.0%
$25,000 – $49,999	29,090	21.7%	29,764	21.9%	0.2%
$50,000 – $74,999	23,375	17.4%	22,210	16.4%	-1.1%
$75,000+	53,160	39.6%	54,922	40.4%	0.8%
1-person	40,735	30.4%	42,666	31.4%	1.1%
2-person	46,000	34.3%	45,870	33.8%	-0.5%
3-person	20,390	15.2%	20,630	15.2%	0.0%
4+person	27,030	20.1%	26,634	19.6%	-0.5%
TOTAL HHS	134,155	100.0%	135,800	100.0%	

Appendix A (continued)
Comparison of 2006-2010 CTPP with 2009-2013 ACS
(Shaded Distribution-Change Cells are at Least 1% in Absolute Value)

J. KATS

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	30,360	27.8%	29,858	27.2%	-0.6%
$25,000 – $49,999	28,272	25.9%	28,153	25.6%	-0.3%
$50,000 – $74,999	19,584	17.9%	19,644	17.9%	-0.1%
$75,000+	30,957	28.4%	32,268	29.4%	1.0%
1-person	32,250	29.5%	32,850	29.9%	0.3%
2-person	37,514	34.4%	38,245	34.8%	0.4%
3-person	16,659	15.3%	16,912	15.4%	0.1%
4+person	22,750	20.8%	21,916	19.9%	-0.9%
TOTAL HHS	109,173	100.0%	109,923	100.0%	

K. WestPlan

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	24,148	27.9%	24,188	27.9%	0.0%
$25,000 – $49,999	24,608	28.5%	24,009	27.7%	-0.7%
$50,000 – $74,999	16,937	19.6%	16,367	18.9%	-0.7%
$75,000+	20,745	24.0%	22,007	25.4%	1.4%
1-person	22,856	26.4%	22,370	25.8%	-0.6%
2-person	31,144	36.0%	32,168	37.2%	1.1%
3-person	13,122	15.2%	13,360	15.4%	0.3%
4+person	19,316	22.3%	18,673	21.6%	-0.8%
TOTAL HHS	86,438	100.0%	86,571	100.0%	

L. JACTS

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	15,965	26.4%	16,276	27.1%	0.7%
$25,000 – $49,999	16,489	27.3%	15,644	26.0%	-1.2%
$50,000 – $74,999	12,108	20.0%	11,939	19.9%	-0.2%
$75,000+	15,913	26.3%	16,241	27.0%	0.7%
1-person	16,034	26.5%	16,297	27.1%	0.6%
2-person	21,799	36.0%	22,352	37.2%	1.1%
3-person	9,114	15.1%	8,883	14.8%	-0.3%
4+person	13,528	22.4%	12,568	20.9%	-1.5%
TOTAL HHS	60,475	100.0%	60,100	100.0%	

Appendix A (continued)
Comparison of 2006-2010 CTPP with 2009-2013 ACS
(Shaded Distribution-Change Cells are at Least 1% in Absolute Value)

M. TwinCATS + NATS

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	17,096	29.9%	16,376	29.6%	-0.3%
$25,000 – $49,999	14,927	26.1%	14,385	26.0%	-0.1%
$50,000 – $74,999	10,248	17.9%	9,747	17.6%	-0.3%
$75,000+	14,934	26.1%	14,861	26.8%	0.7%
1-person	16,606	29.0%	16,571	29.9%	0.9%
2-person	20,385	35.6%	20,227	36.5%	0.9%
3-person	8,037	14.0%	7,745	14.0%	-0.1%
4+person	12,177	21.3%	10,826	19.6%	-1.7%
TOTAL HHS	57,205	100.0%	55,369	100.0%	

N. MACC

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	8,218	18.8%	8,205	18.8%	0.0%
$25,000 – $49,999	12,470	28.6%	11,761	27.0%	-1.6%
$50,000 – $74,999	9,640	22.1%	9,056	20.8%	-1.3%
$75,000+	13,274	30.4%	14,602	33.5%	3.0%
1-person	9,432	21.6%	8,992	20.6%	-1.0%
2-person	14,805	34.0%	15,031	34.5%	0.5%
3-person	6,580	15.1%	6,915	15.9%	0.8%
4+person	12,785	29.3%	12,686	29.1%	-0.2%
TOTAL HHS	43,602	100.0%	43,624	100.0%	

O. BCATS

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	10,755	28.7%	10,908	29.3%	0.6%
$25,000 – $49,999	10,917	29.1%	10,798	29.0%	-0.1%
$50,000 – $74,999	7,036	18.8%	6,639	17.8%	-1.0%
$75,000+	8,766	23.4%	8,904	23.9%	0.5%
1-person	11,377	30.4%	11,661	31.3%	0.9%
2-person	12,613	33.7%	12,586	33.8%	0.1%
3-person	5,855	15.6%	5,711	15.3%	-0.3%
4+person	7,629	20.4%	7,291	19.6%	-0.8%
TOTAL HHS	37,474	100.0%	37,249	100.0%	

Appendix A (continued)
Comparison of 2006-2010 CTPP with 2009-2013 ACS
(Shaded Distribution-Change Cells are at Least 1% in Absolute Value)

P. TVC

Category	2006-2010 CTPP		2009-2013 ACS		Distribution Change (%)
	Household	Distribution (%)	Household	Distribution (%)	
< $25,000	6,700	20.0%	7,011	20.9%	1.0%
$25,000 – $49,999	9,426	28.1%	8,692	25.9%	-2.1%
$50,000 – $74,999	7,290	21.7%	7,067	21.1%	-0.6%
$75,000+	10,162	30.3%	10,742	32.1%	1.8%
1-person	9,304	27.7%	9,473	28.3%	0.6%
2-person	13,512	40.2%	13,073	39.0%	-1.2%
3-person	4,237	12.6%	4,878	14.6%	1.9%
4+person	6,525	19.4%	6,088	18.2%	-1.3%
TOTAL HHS	33,578	100.0%	33,512	100.0%	

Appendix EEE

Data Integration GPS Correction Factor Plan

GPS Correction Factors

Background

It is widely acknowledged that respondents, when self-reporting or reporting for others in their household, typically under-report trips in household travel surveys. One common method to measure and correct for trip under-reporting is to deploy GPS data loggers to a sample of households and compare the self-reported trips to those collected passively by the GPS loggers. Analysis is conducted to identify circumstances that lead to under-reporting, and correction factors are calculated. Because there is substantial variation that occurs in under-reporting, some households may not require a correction factor whereas others will require a fairly large weight (see Zmud & Wolf, 2003[1]); applying a single correction factor to all log-reported travel is not appropriate. There are several factors or conditions that can contribute to trip under-reporting. For example, trip length has been shown to be correlated with trip under-reporting. Trips of short duration are often missing from respondent logs more frequently than trips of long durations (Zmud & Wolf, 2003). Variables examined in the past have included trip duration, household size, reported vehicle ownership, household income, respondent age, employment status, student status, and presence of children under 18 in the household.

As noted above, the traditional travel correction factor generation process applies to study designs where GPS and survey data are collected concurrently and independently in what is often referred to as the *dual-method* design. This approach does not apply to the MTC III and STC15 prompted recall design, in which the GPS data serve as a basis for travel reporting and confirmation, but we believe it is helpful to explain the traditional approach Westat used in previous studies so that differences in the two approaches are known and understood.

Under the dual-method design, the process for developing travel correction factors begins with a database of GPS trip records used to test a model of trip misreporting. A dummy variable is used to indicate if a trip record was "missing" when compared to the GPS data. A logistic regression model is then used to determine which of the variables associated with the trip record (e.g., household size, household income, employment status, age, etc.) has a statistically significant effect on travel under-reporting. Based on the logistic regression analysis, adjustment factors are developed for specific combinations of household, person, and trip types. These factors are then applied to all households, allowing for more accurate estimation of household trip rates for a wider population.

The process for developing correction factors for MTC III and STC15 is notably different from the traditional trip correction process because of the project's use of the GPS prompted recall (GPS-PR) design. Instead of comparing two 'streams' of data from the same sub-sample of households to identify underreported travel, this process involves a comparison of GPS-captured travel made and confirmed by GPS households with the reported travel behavior equivalents from non-GPS households. The thought is that the participants using the GPS-PR retrieval

[1] Zmud, J., and J. Wolf. (2003). Identifying the correlates of trip misreporting - Results from the California Statewide Household Travel Survey GPS Study. In *Proceedings of the 10th International Conference on Travel Behaviour Research*, IATBR, Lucerne, Switzerland.

method are more likely to capture all travel movements on the travel day. Under this assumption, data from the GPS households are used to identify and quantify travel under-reporting in non-GPS households, and estimate correction factors that can be applied to more fully account for the travel of the full sample.

Development process

Westat applied a modified version of the methodology developed and executed by Parsons Brinckerhoff (PB) for use on the New York Metropolitan Transportation Council (NYMTC) Regional Travel Survey (RTS[2]). This method applied adjustments at the tour level based on tour and tour stop frequency differences between the non-GPS and GPS groups. A tour consists of all the travel reported from when a participant left their home to when they returned. Since a tour consists of multiple places, determined by the primary destination (e.g., work place) and other stops made on the tour, these correction factors can be applied to tours and places when analyzing the travel data. Another important component of the method developed for NYMTC was grouping the tours by person types typically used in Activity Based Models (ABMs).

Similar to the PB method, Westat structured the GPS correction derivation process around tour and tour stop frequencies. However, Westat did not use tour generation models to determine statistical significance of travel under-reporting in the non-GPS portion of the sample. Instead of following this approach, a process tailored to the MTC III and STC15 was developed to compute travel correction factors (TCFs). It included the following general steps:

1. Prepare tour data files that translate place-based survey database to tours, as well as linked trip files; these allow the estimation of tour and tour stop frequencies.
2. Identify tour segmentations (e.g., tour type, person type, and tour mode combinations) for which statistically significant tour frequency distribution differences exist between GPS and non-GPS groups.
3. Use a similar exploratory approach to identify tour stop frequency segmentations (e.g., by tour mode and stop counts) that show under-reporting in the non-GPS group.
4. Utilize person-level weights to expand the GPS and non-GPS tour frequencies for the segmentations found to have significant under-reporting and compute tour frequency TCFs (i.e. compute weighted tour frequencies using the person level weights for GPS and non-GPS and use them to calculate a ratio).
5. Compute tour stop frequency TCFs by tabulating tours using person weights multiplied by the tour frequency travel correction factors derived in step 4 (i.e., compute weighted tour frequencies classified by number of stops using the product of person level weights and tour frequency TCFs for GPS and non-GPS groups and use them to calculate a ratio).

[2] Full final report available at https://www.nymtc.org/portals/0/pdf/RHTS/RHTS_FinalReport%2010.6.2014.pdf.

The process of tabulating the tour data, applying the tests, and computing TCF values was automated using the R statistical package[3]. Before the tour frequency distributions were compared, they were capped at three tours per person; that way the categories for aggregating the tour frequency distributions at the person level were: zero, one, two, and three-plus. A minimum of six observations on either group (GPS and non-GPS) was stipulated for performing the tests and considering the computation of TCFs.

Simple tour frequency multinomial logit models and the Kruskal-Wallis[4] (KW) non-parametric test were used to identify significant differences in tour and tour stop frequencies between the GPS and non-GPS groups. The null hypothesis for the KW test is that the compared samples are from identical populations (i.e., the tour frequency distributions do not differ between GPS and non-GPS groups). This test was selected given the heavily skewed nature of the tour frequency distribution, which did not make it possible to apply parametric tests such as Analysis of Variance (ANOVA). The simple tour frequency multinomial logistic regression models had the tour frequency as the dependent variable and had only a dummy variable indicating whether the person was GPS instrumented as an independent variable (or covariate). The resulting GPS coefficient's p-value was used in combination with the KW test statistic to help assess significance of the GPS as a factor in explaining tour frequency differences for the various segmentations examined. A significance level of at least 10% was required for a segmentation to be considered. A similar approach was adopted when comparing tour stop frequencies between GPS and non-GPS tour data, including the 10% significance level. The stop frequencies were capped at three, such that tests were performed only for stop frequency values equal to one, two, and three-plus stops.

Applying Corrections

The computed TCFs should be used whenever summarizing travel attributes, such as traveled distance in conjunction with existing household or person weights. For example, to compute average trip distance by person, one should multiply individual place distances by the product of the person weights and the place TCFs, sum the resulting values and then divide by the sum of the product of the person weights and TCFs as noted in the following equation:

$$\bar{d} = \frac{\sum d \cdot w_p \cdot TCF_{freq} \cdot TCF_{stop}}{\sum w_p \cdot TCF_{freq} \cdot TCF_{stop}}$$

where:

d	:	trip distance,
w_p	:	person level weight,
TCF_{freq}	:	tour frequency travel correction factor, and
TCF_{stop}	:	stop frequency travel correction factor.

[3] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[4] Myles Hollander and Douglas A. Wolfe (1973), *Nonparametric Statistical Methods.* New York: John Wiley & Sons. Pages 115–120.

Summary of Correction Factors

The process outlined in the previous section was applied separately to the MTC III and STC15 datasets. The tour data were merged with household and person datasets such that only those households present in the final deliverable files were included.

MTC III GPS Travel Correction Factors

When reviewing the MTC III tour data, the following segmentations were determined to show statistically significant travel under-reporting:

- Tours made by workers and students
- Tours made by households with children
- Tour stops under 15 minutes of duration

Other segmentations that were analyzed, but found to be insignificant (requiring no adjustment), were household size, tour purpose, tour mode, and tour duration.

Tour rates were calculated for the GPS and non-GPS groups by segments. A tour correction factor was calculated for each segment as the ratio of the tour rates for GPS to non-GPS groups (minimum ratio is constrained to one). The tour rates were computed using the person level weights. The non-GPS tours were further corrected to account for the fact that the tour rates for tours with a higher number of stops are higher for the GPS instrumented participants. Tours with stop durations shorter than 15 minutes were identified to show significant differences between the GPS and non-GPS groups. For these tours, stop frequency travel correction factors were calculated as the ratio of tour rates between GPS and non-GPS groups using the person weights multiplied by the tour frequency travel correction factors.

Tables 1 and 2 show final TCF values for the segments found to have significant under-reporting. Cases where the computed tour frequency ratios were less than 1.00 were not included in the final correction factors attached to the tour and place data.

Table 1 – Tour Frequency Correction Factors by Presence of Children

Person Type	No Children in Household	Children in Household
Worker	1.08	1.17
University Student	1.04	1.28
Non-Worker	NS*	NS*
Retiree	NS*	NS*
Driving Age School Child	NA**	1.12

* Tour frequency found not to be significant.
** Not enough records to compute a rate.

Table 2 – Stop Correction Factor by Number of Short Stops (Duration < 15 minutes)

Person Type	One Stop	Two Stops	Three Stops
All	1.06	1.08	1.02

STC15 GPS Travel Correction Factors

Similar to what was found when reviewing the MTC III tour data, the following tour frequency segmentations were determined to show statistically significant travel under-reporting for the STC15 dataset:

- Tours made by workers and retirees
- Tours made by households with children
- Tour stops under 15 minutes of duration

For the STC15 dataset, tour stop frequency was not found to show as much under-reporting as was observed in the MTC III dataset, with lower significance overall. Tables 3 and 4 contain the computed correction factors for the STC15 tours.

Table 3 – Tour Frequency Correction Factors by Presence of Children

Person Type	No Children in Household	Children in Household
Worker	1.05	1.16
University Student	NS*	NS*
Non-Worker	NS*	NS*
Retiree	1.09	NS*
Driving Age School Child	NA**	1.12

* Tour frequency found not to be significant.
** Not enough records to compute a rate.

Table 4 – Stop Correction Factor by Number of Short Stops (Duration < 15 minutes)

Person Type	One Stop	Two Stops	Three Stops
All	NS*	NS*	1.02

* Tour stop frequency not found to be significant.