



MEASURES FOR SUCCESS:
New Tools for Shaping
Transportation Behavior
July, 2017

CONTENTS

Acknowledgements.....3

Executive Summary.....4

Introduction.....6

Individualized Marketing: What It Is.....6

The State of Evaluation.....8

Protocols.....11

Protocol #1. Logic Models for Evaluation.....12

Protocol #2. Segmentation Approach / Change in Psychological Stage.....15

Protocol #3. Survey Design and Analysis.....18

Protocol # 4. Passive Tracking.....24

Protocol # 5. New Metrics.....26

Protocol # 6. MAX-SUMO Framework.....32

Protocol #7. A/B Testing.....34

Five Recommendations for Better Behavior Change Programs.....37

References and appendices available online at transitcenter.org

ACKNOWLEDGEMENTS

TransitCenter gratefully acknowledges the team at Alta Planning + Design that produced this report, particularly Jessica Roberts, Derek Hofbauer, and Katie Mangle. TransitCenter's Steven Higashide and Kirk Hovenkotter oversaw the research and provided additional contributions.

TransitCenter and Alta Planning + Design thank the following individuals and organizations for their contribution to this project:

Elizabeth Ampt – Concepts of Change
 Josh Berezin – Analyst Institute
 Megan Kennedy-Chouane, Camille Fabbri – Children's Investment Fund Foundation
 Dr. Jana Corbett – independent consultant
 Alexandra DeMoss, Paul Mason, Janice Helman – Washington State Department of Transportation
 Lori Diggins – LDA Consulting
 Dr. Jennifer Dill – Portland State University
 Elizabeth Green – Resource Systems Group, Inc.
 Philip Li – Robert Sterling Clark Foundation
 Elizabeth Linos – Behavioral Insights Team
 Todd Litman – Victoria Transport Policy Institute
 Craig Nelson – Steer Davies Gleave
 Eric Schreffler – Eric Schreffler Transportation Consultant
 Jeff Tyrens – former Director of Oregon Progress Board
 Caleb Winter – Portland Metro Regional Government
 Duncan Wyse – Oregon Business Council
 Dr. Adrienne Zell, Dr. Kerri Winters-Stone – Oregon Health and Science University

Design: Hayley Richardson
 Published by: TransitCenter
 1 Whitehall Street
 New York, NY 10004
www.transitcenter.org

(C) 2017 TransitCenter

EXECUTIVE SUMMARY

If you build it, will they come? As more and more American cities invest in transit, bikeshare, and biking and walking facilities, this is the key question. If citizens use new transportation options in large enough number, they can help tackle some of the biggest challenges cities face: Managing traffic, improving public health and affordability, and reducing pollution and greenhouse gas emissions.

While use of transit and active transportation is highly dependent on the quality of infrastructure and service provided, significant numbers of people can be converted from potential to actual users of sustainable transportation through the use of marketing and behavior change programs.

One of the most successful types of transportation behavioral change programs is known as individualized marketing (IM). Over the last two decades, dozens of cities in the US have successfully used IM campaigns to make people aware of transit, biking, and walking options in their neighborhood, persuade them to try those options, and ultimately convince them to use those options long-term. These campaigns reach out to all households in a target neighborhood and identify residents who are interested in learning more about alternatives to driving. Interested participants can receive coaching, information, events, communications, and materials. In contrast to traditional “blanket” marketing, IM programs are cost-effective because they engage motivated participants, and provide them with targeted information that is specific to their neighborhood. Cumulatively, these programs have helped hundreds of thousands of people to try transit, walking, bicycling, and shared mobility—and eventually make permanent changes to their transportation habits.

IM has always had an ethos of strong evaluation and reporting practices. However, practitioners know that IM evaluation hasn’t evolved enough in the last two decades. This means opportunities are being missed to tell new stories about how changing transportation habits can transform lives, form new partnerships with funders and policymakers, and make programs more efficient.

This report provides new tools and ways of thinking for the individualized marketing professional in mind. However, the ideas presented have potential value for any transportation behavior change effort.

Interviews conducted with evaluation experts and academics were instrumental in developing appropriate protocols (e.g. methods and tools) for this report to help transportation demand management (TDM) practitioners better evaluate programs. The project team then conducted a literature review to document current research, lessons learned, and case studies from fields such as behavioral economics, health, and statistics. A 10-member Advisory Committee comprising IM practitioners from around the US met four times during one year to discuss and refine the research findings and protocol development process.

This report contains seven new evaluation protocols for use by TDM professionals that were developed through the research and guidance conducted with these experts and practitioners. In this report, the term “protocol” is used as an umbrella term that includes techniques, tools, and new ideas that may help IM practitioners in their evaluation. Instructions, guidance and additional resources are provided in each protocol. The protocols presented in this report offer a menu of options, to be used alone or in combination, depending on the needs of the program.

1. Logic Models
2. Segmentation Approach/Change in Psychological State
3. Survey Design and Analysis
4. Passive Tracking
5. New Metrics
6. MAX-SUMO
7. A/B Testing

TDM professionals can use the protocols presented in this report to develop more robust evaluation frameworks and improve existing survey design and analyses methods. By doing so, they can better understand and present program results, and satisfy the concerns of funders and decisionmakers.

PROJECT ADVISORY COUNCIL

Tien-Tien Chan, City of Austin, Texas

Carol Cooper, King County Metro

Heleen Dewey, Spokane Regional Health District

Linda Ginenthal, Portland Bureau of Transportation

Sarah Goforth, Portland Bureau of Transportation

Alex Keating, New York City Department of Transportation

Antoinette Meier, San Diego Association of Governments

Emma Pachuta, St. Paul Smart Trips

Krute Singa, San Francisco Department of Environment

Sean Wiedel, Chicago Department of Transportation

INTRODUCTION

Well-managed transportation systems can help cities meet goals for improving air quality, reducing emissions, and improving quality of life. In order to get there, cities need to complement investments in public transit with policies that remove hidden biases towards the car, coupled with programs that effectively market and promote transit, biking, and walking. All of these – capital and service investments, policies, and programs – are often classified under the umbrella of transportation demand management (TDM). Traditional employer-based TDM programs reach employees in the work setting and focus primarily on commute trips. Yet only one in five trips is a commute trip. Individualized Marketing campaigns are one way to reach all residents – not just employees – and address all trip types and modes – not just the commute to work. Time-limited IM campaigns offer some distinct opportunities for evaluation – most notably, because they are limited in time and in space, they allow the short-term impacts of the campaign to be measured and expressed with greater confidence.

INDIVIDUALIZED MARKETING: WHAT IT IS

What is individualized marketing? Imagine that Diana lives in a neighborhood that has been selected for an IM campaign. She drives most places out of habit, but sometimes thinks it would be nice to walk or bike to the library for exercise, or get some knitting done on the bus. She hasn't taken any action to figure out how to do those things.

Diana receives a mailing about a new program just for her neighborhood. She's given the chance to order a customized travel information packet of materials she chooses – and she can select a small gift (such as a 5-pack of transit tickets or a bike light) to go with it. She's pleased that the packet arrives at her door within a week, and she spends some time with the maps and materials.

A week later, Diana is contacted by a travel ambassador, asking if she'd like some help putting her information to use, and she admits that she'd like to try the new bike share system, but she feels intimidated by the signup process. The travel ambassador arranges to meet her at a bike share station with a free 24-hour pass; he helps her sign up for the system on the spot and takes her on a short bike ride of a bike route to the library near her house.

Diana receives other communications from the program – a mailed newsletter, some follow-up emails, and Facebook posts. Through these, she learns that there's a new pedestrian crossing being installed at the light rail stop near her house. She also learns some helpful tips about transportation options, and she reads features of some of her neighbors talking about the benefits they get from traveling by bus, foot, and transit.

Diana decides to join up with a walking tour of public art and community gardens, hosted by the program. It meets up at a nearby park, and on the walk, she makes an acquaintance who also wants to walk more; they decide to organize a neighborhood weekly walking club.

By the time the program has wrapped up four months later, Diana has much more information about using transit, bikeshare, and walking. Plus, she has had the opportunity to try out these modes in a friendly, supportive, and social atmosphere. Now she's better equipped to leave the car at home when it isn't the best choice for her.

Diana doesn't know it, but she has just benefitted from an IM program. The program was organized by her city's Department of Transportation; at the end of the program their evaluation shows that program participants are making 10 percent fewer solo car trips than they did before the program began.

This fictional program is typical of IM programs, which promote transportation options to people within a geographic target area and/or to a demographic target audience. Signature aspects of IM campaigns include a focus on all trips (not just commute trips) and all sustainable travel modes; a time-limited campaign (usually between three and six months); and an opt-in model that invites everyone to participate but focuses resources on people who self-identify as interested. IM campaigns assume that the primary barriers to using non-Single Occupant Vehicle (SOV) modes are low awareness of options, low confidence about using options, and strong habits around driving.

Tabling at Farmer's Markets is a great way to provide transportation options materials and information to interested residents



Real-life IM programs can vary significantly from the generic model. They may include trip logging; they may not include events; the personalized coaching may play a larger or smaller role. The large family of IM programs includes existing programs known as SmartTrips, SmartTravel, TravelSmart, personalized travel planning, or Travel Blending. IM programs have been successfully implemented in Europe, Australia, Japan, Canada, and the United States.

IM campaigns offer some distinct opportunities for evaluation—short-term impacts of the campaign can be measured and expressed with greater confidence compared to standing programs that do not have a start or end date. A 2015 European Union review of sustainable transportation strategies found extensive evidence that IM programs shift travel behavior, at least in the short-term. In the U.S., the most robust evidence comes from Portland, Oregon, which has conducted large-scale individualized marketing programs for over a decade. Portland's campaigns have typically targeted 20,000 to 40,000 households each year, of whom roughly a quarter ultimately participate in the program. Evaluations of those programs have consistently found meaningful mode shift following program implementation in the range of 9 percent relative reduction in drive-alone trips.

Like all TDM programs, IM is more successful when there is supportive transportation infrastructure. There is not much point in marketing a bad product; and in this case, the product is transit, walking, and biking. Several studies have shown that reductions from IM programs persist longer in neighborhoods that are better for walking, bicycling, and transit.

THE STATE OF EVALUATION

While many US cities have run IM campaigns, nearly all of them have been at a much smaller scale than the programs run in Portland, Australia, or the United Kingdom, where the program may serve tens of thousands of residents at a time.¹ For example, the Oregon Department of Transportation (ODOT) recently funded eight IM programs across the state; however, the sample of households targeted was approximately 5,000 for each program. Many other cities across the US have also invested in IM programs, but serve thousands of residents per program rather than tens of thousands.

Successful IM campaigns have been conducted in the US over the past decade. Compared to many TDM efforts, IM campaigns have a strong commitment to evaluation, which has led to a reputation for reliable results. At the same time, the small size of the typical campaign (and small budget) has afforded little opportunity for IM practitioners to research and try new evaluation methods.

Today, most IM campaigns rely on some form of self-reported survey data to determine if any mode shift has happened. There are known concerns with using self-reported survey data, related both to bias in who takes the survey, and bias and error in what data gets reported. Even aside from those concerns, a one-day snapshot of travel behavior may be influenced strongly by that day's weather, yet a general survey (e.g. "what percentage of the time do you travel by the following modes?") is considered less accurate. This report offers both alternatives to surveys and methods to improve the quality of survey data.

Secondarily, measuring vehicle miles reduced (VMR) as the main way to measure a program's impact is too limiting. For example, someone who has begun walking to work one day a week has made a significant change in their behavior – yet a one-day trip diary may not capture that change. Likewise, a VMR-only approach misses out on opportunities to talk about other kinds of transportation-related successes: someone who has gone from being a timid to a confident user of travel options; someone who got their bicycle in working order and tried it out for the first time in years; someone who is still driving much of the time, but is much more open to and positive about travel options; someone who sold a car as a result of the program. Beyond mode shift, there are opportunities to daylight benefits that are currently not measured, such as those related to economic development, health benefits, and social cohesion/community-building.

That's why TransitCenter commissioned Alta Planning + Design to explore whether practitioners are using the most current tools and approaches for influencing and measuring behavior change, and what "next-generation" approaches to evaluation the field should adopt.

Advised by public-sector practitioners, Alta brought in methods from TDM evaluation experts and from evaluation specialists in other fields. The result is seven evaluation protocols that have value to IM practitioners and to the larger behavior change and TDM community. They can be used separately or in combination, and they do not require specialized training or software to implement. Ultimately, the goal is to build a larger toolbox of evaluation options that will help practitioners speak confidently about the impact of their work, make the case for behavior change programs, and improve campaigns over time.

ABOUT THE RESEARCH

As the first step of the project, the team researched the state of the practice for evaluating IM programs. Next, twenty-one evaluation experts representing seventeen different organizations were selected to take part in a one-hour phone interview with TransitCenter and the Alta research team. The interviews covered how those disciplines approach evaluation, and what tools and innovations they use. Interviewees came from fields including public health, TDM evaluation, travel behavior research, statistics, economic policy organizations, climate change, behavioral economics, political research, and regional government.

Key interview findings indicate:

- Evaluation methods for IM programs vary among both practitioners and researchers.
- New metrics for IM programs can be established without major changes to project design.
- Technology is reducing the costs of data collection and evaluation.
- Theoretical frameworks show strong potential for providing more consistent and defensible project results.
- There is increasing demand by funders for accountability and rigorous evaluation.

Additionally, the Alta team conducted a thorough literature review, which includes the following major takeaways:

- IM programs typically rely on small data sets for analysis; it is important to increase the scale of these programs to provide more meaningful and statistically significant analysis.
- Standardizing survey questions allows IM programs to be more comparable over time.
- Control groups can help isolate the effects of external factors such as weather or new infrastructure.
- New technologies have the potential to change how travel behavior data is collected.
- Evaluation timeframes differ between IM programs, and sustainability of results has been shown to last between one and three years.²
- Methods commonly used in public health, including the Transtheoretical Model and logic models, could be applied successfully to IM campaigns.

The interview and the literature review findings, which are presented in the Research Summary Report, were instrumental for developing the seven evaluation protocols presented in this final report.

PROTOCOLS

Protocols are intended to provide methods and tools that help TDM practitioners evaluate programs in new, expanded, or different ways. For each protocol, we describe what is proposed and explain when and why to use it. Instructions, guidance, and additional resources are provided to make it easier for practitioners to implement. The protocols in this report are intended to offer a robust set of evaluation tools and ideas to TDM practitioners. They will help to answer questions such as:

- Is this program reducing drive-alone trips and vehicle miles traveled?
- Is the selected evaluation approach based on best practices? Will it withstand criticism?
- What types of program benefits to communities and individuals can we reveal?
- How can a given program be made more effective?

This research program was specifically designed around the needs of IM campaigns, which have some features that make standard TDM evaluation approaches challenging. Most notably, they do not focus on the commute trip and they address all trips and all modes in a given area (usually a residential neighborhood). There is a great need for innovation and information sharing in the IM practitioner community; therefore, the protocols listed in this report will provide practitioners with the necessary guidance and instructions to integrate new evaluation strategies into both new and existing IM programs.

A campus IM participant pledging to drive less



PROTOCOL #1. LOGIC MODELS FOR EVALUATION

The Takeaway: Logic models are a tool for mapping out assumptions about how a program works. The process of building a logic model can help practitioners identify the program elements that should be evaluated. Logic models map out, in a diagram form, assumptions about how an intervention works. Logic models list what resources are available, how they will be used, and the intended result (outputs and outcomes). Using logic models in program design and evaluation is quite common in public health, but is less common for transportation behavior change campaigns.

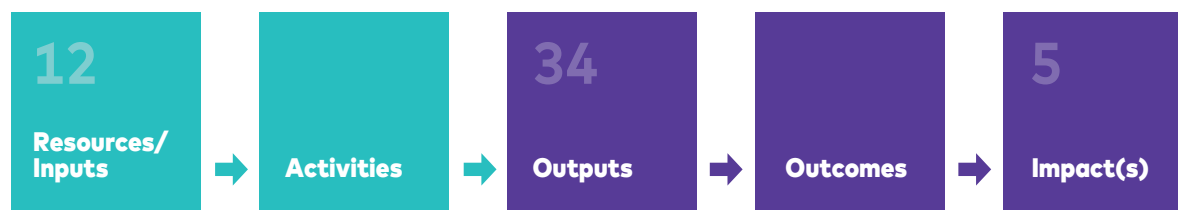
The process of developing a logic model can help to identify key assumptions that the evaluation should test. For example, if the program designer assumes that distributing customized travel packets leads to behavior change, data should be collected both on the number of packets distributed (outputs) as well as whether participants report that the packets helped them change their behavior (outcomes). Bringing this practice into TDM campaigns may raise the level of rigor in evaluation design.

Because the use of logic models creates transparency around how a program changes behavior, it can have additional benefits. Funders (especially those without experience in transportation) often appreciate the use of logic models because it helps them understand why the program works. The process of creating a logic model can also lead to better program design, as practitioners see opportunities for better meeting program goals or taking advantage of resources.

Logic models can be simple or complex. They can vary in their design, the categories shown, and the complexity of the causal relationships. The following steps will result in a logic model suitable for most IM campaigns.

See recommended online resources listed in the footnotes for more information and guidance.

Figure 1: Logic model steps



Your Planned Work

Your Intended Results

Source image: W.K. Kellogg Foundation Logic Model Development Guide." January 2004.
<http://www.smartgivers.org/uploads/logicmodelguidepdf.pdf>

Step-by-Step Methodology

Each component of the logic model should be connected to the others through a causal connection. When the logic model is complete, test whether a connection can be made for each element. If not, identify gaps and adjust the model. Some people find it useful to start with impacts and work backwards to resources, while others prefer to work forwards from resources and end with impacts.

Step 1: Resources

- List the resources (also called inputs or assets) that are available to support your program.³
- Examples: full or part-time staff, consultants, grant funding, program materials.

Step 2: Activities

- Identify the activities/actions that are needed to implement your program. Activities should make use of resources in order to achieve the desired results.
- Examples: distributing information (e.g. promotional or educational materials); providing services (e.g. education, counseling, or events).

Step 3: Outputs

- For each of the specific activities that you have listed, identify the outputs (measurable, tangible products or results of program activities) you hope to achieve through your program.
- Examples: number of workshops put on, hours of service provided, marketing materials developed, and number of partnerships formed.

Step 4: Outcomes

- Identify the outcomes you expect to realize through the specific activities undertaken. For IM programs, this should focus on the change anticipated for program participants. If desired, outcomes can be further separated into short-term and intermediate outcomes.
- Examples: increased awareness of a program (short-term); 30 percent more children, ages eight to eleven, that are walking and biking to school over five years (intermediate)
- Outcomes should be SMART: Specific, Measurable, Action-oriented, Realistic, and Timed.⁴

Step 5: Impacts

- Long-term outcomes/impacts: The long-term impact or systemic change you expect your program to achieve. For an IM program, this should be focused on system impacts at the neighborhood or community level.

- Examples: increase in percentage of population achieving recommended daily physical activity thresholds; decrease in drive-alone mode share; increase in transit ridership; decrease in measured pollutants in the air.⁵

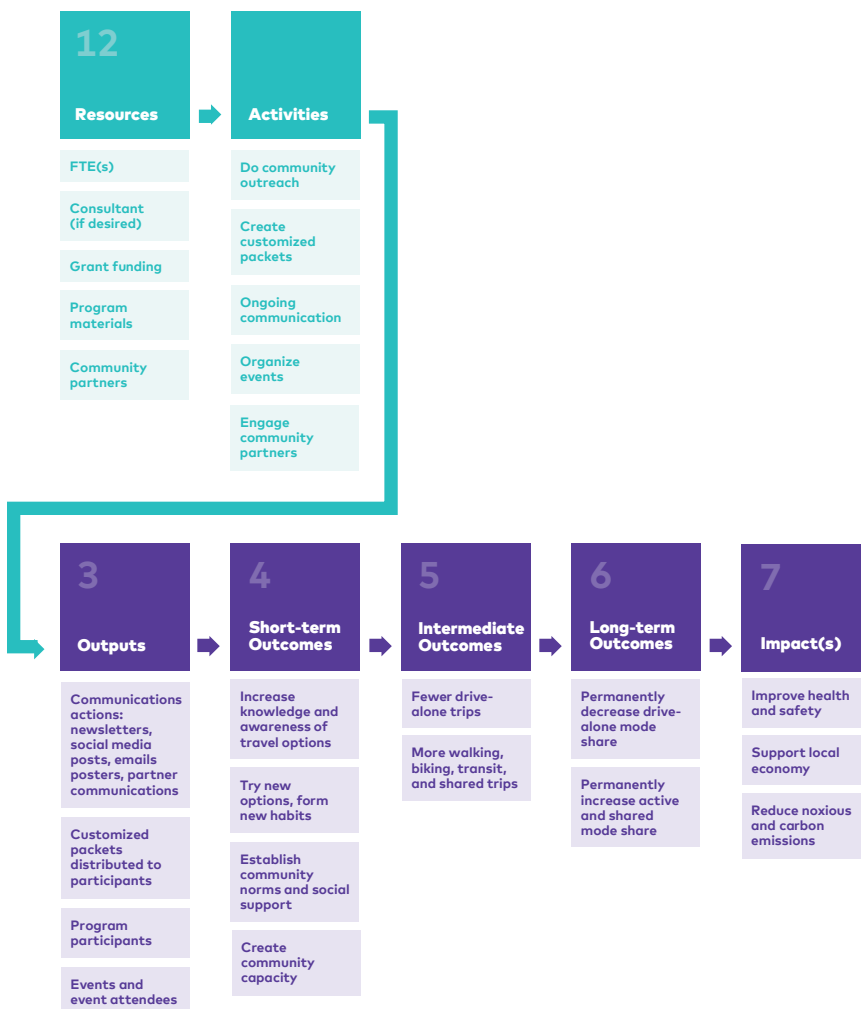
Step 6: Make Connections

Once you have identified the components of your model, the next step is to put it in graphic form, adding arrows that show relationships between each of the components. Simple logic models only add a single arrow between each component (e.g. a single arrow that connects all the resources to all the actions; see Figure 2 for an example). Complex logic models may include arrows that show a causal pathway between specific individual components (e.g. showing one particular output leading to one particular outcome). Choose the level of complexity that best suits the type of program.

Step 7: Solicit Feedback and Refine

When you have finished a draft of the model, ask others to review it for accuracy and readability. Refine and revise it until both you and others who have provided feedback are satisfied.

Figure 2: Logic Model Example (Source: Alta Planning + Design)



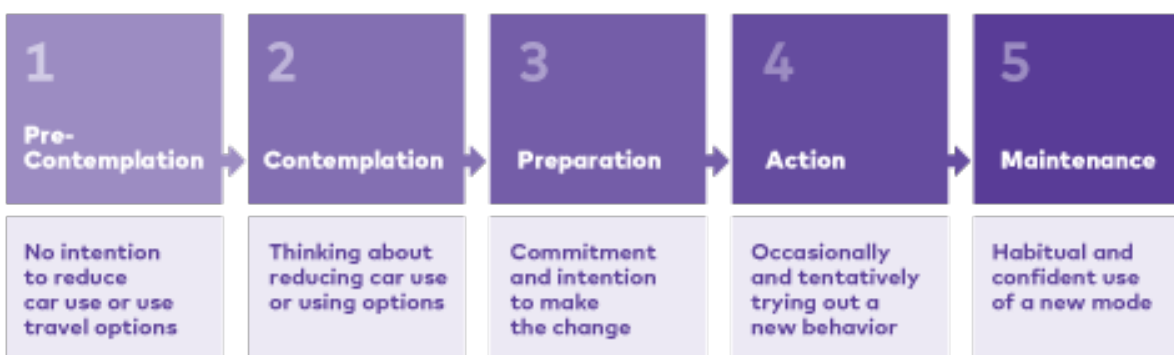
PROTOCOL #2. SEGMENTATION APPROACH / CHANGE IN PSYCHOLOGICAL STAGE

The Takeaway: Assessing someone's readiness to make a behavior change (also called their "stage of change") can help to provide tools and coaching that are more appropriately targeted. Observing whether a program has resulted in a change in Stages of Change assessments can reveal program impacts that are not apparent in a traditional travel diary evaluation.

The Stages of Change model (also called the Transtheoretical Model) is a well-accepted framework for understanding how behavior change happens. Program participants are categorized based on how willing and interested they are in changing their own behavior. Fundamentally, the Stages of Change model suggests that only people who are interested in changing their behavior will successfully make that change. The model also suggests that the interventions (e.g. messages and activities) that are appropriate for someone in one stage are not effective for someone in another stage.

A program based on Stages of Change begins by sorting program participants into stages; this is called segmentation. Then, evidence-based strategies are offered to participants based on their stage. The purpose of stage-specific interventions is to move participants to a more advanced stage of change, culminating in the permanent desired behavior change. Often used in addiction treatment programs, this theory can be useful during program design by allowing participants to receive evidence-based interventions appropriate to their current stage of change. This model has also been used successfully for transportation behavior change programs in Europe. The overarching goal of applying the Stages of Change model is to create more effective and more defensible programs.

Figure 3: Stages of Change Segmentation



From an evaluation perspective, using Stages of Change segmentation can lead to a new way to understand and discuss success. If people are moving from one stage to the next, that can be an indication that the program is successful. This approach may allow practitioners to make visible progress that is not revealed by a traditional mode-shift evaluation. For example, someone who has progressed from Contemplation to Action may not be performing the new action often enough for it to show up on a trip diary, yet significant behavior (and psychological) change has occurred. Likewise, if someone has moved from Action (occasionally/tentatively trying out a new behavior) to Maintenance (consistent, habitual, confident use of the new mode), that represents progress and can be called out in an evaluation – yet may not be revealed through trip diaries. If many respondents progress from Contemplation to Action, but not to Maintenance, this may indicate that one aspect of the program or transportation system isn't working as well as the rest. (For example, an effective campaign may convince many people to try transit, but if the transit service is not useful enough for them to maintain the behavior, they may never get to the Maintenance stage.)

For TDM programs, segmentation can either focus on attitude toward car use or attitude toward non-SOV travel options. There is insufficient peer-reviewed research to allow us to recommend one over another. More research is needed to determine whether attitude toward car use is the inverse of attitude toward non-SOV travel options. It may be that they are simply different ways of evaluating the same situation, or it may be that they are in fact slightly different. For example, perhaps a person can be unenthusiastic about reducing their car use (Pre-contemplation) yet simultaneously eager to try out bicycling (Preparation).

For this report, we are recommending two options for segmentation. Practitioners should test both and share results and information with other professionals, with the goal of deciding on a consistent segmentation approach or approaches for IM programs.

Segmentation option 1: Attitude towards transportation options

Select the statement that best applies to you:

- I do not walk/bike/use transit for most of my trips, and I don't intend to make any changes. **[Pre-contemplation]**
- I do not walk/bike/use transit for most of my trips. I am considering walking/biking more often, but I'm not sure how to make that change. **[Contemplation]**
- I do not walk/bike/use transit for most of my trips, but I intend to walk/bike more often. I know how I want to do this, but I haven't yet gotten started. **[Preparation]**
- I walk/bike/use transit to get around for most trips, and I have been doing so for less than six months. **[Action]**
- I walk/bike/use transit to get around for most trips, and I have been doing so for more than six months. ⁶ **[Maintenance]**

Segmentation option 2: Attitude towards car use

Select the statement that best applies to you:

- I drive for most of my trips, and I don't intend to make any changes. **[Pre-contemplation]**
- I drive for most of my trips. I am considering driving less often, but I'm not sure how to do it. **[Contemplation]**
- I drive for most of my trips, but I intend to drive less often. I know what I would like to do instead, but I haven't yet gotten started. **[Preparation]**
- I already use ways other than driving to get around for most trips, and I have been doing so for less than six months. **[Action]**
- I already use ways other than driving to get around for most trips, and I have been doing so for more than six months. **[Maintenance]**

Segmentation should be done during the intake process. For many IM programs, this would be the order form process. For other programs (such as door-to-door), this may be done in person.

IM STAGES OF CHANGE CASE STUDY

Results from a baseline intake (order) form and an online participant-based post survey from an IM program conducted in Milwaukie, Oregon, indicated there was a behavioral shift among the same participants in the panel sample after they received program materials and attended events. Results showed a 16 percent increase in the maintenance phase ("Yes, and it's easy for me") during the post survey evaluation. Participants segmented across different Stages of Change categories received essentially the same program services; however, the results show that offering information and events as part of an IM program increases the likelihood that participants will move further along the behavior change continuum. Offering tailored interventions for each of these groups after the initial segmentation would likely have a greater effect on creating new and lasting behavior changes.

Using segmentation is most appropriate when one is designing a program around the Stages of Change model. It is not appropriate if the program is designed around a different behavior change theory (e.g. Theory of Planned Behavior). Practitioners can use a logic model to show how the segmentation affects program design and delivery; this can reveal evaluation needs. Stages of Change segmentation questions are also a major component of the MAX-SUMO framework, which is referenced in Protocol # 6.

Protocol #3. Survey Design and Analysis

The Takeaway: If surveys will be used in a program evaluation, they should be designed well to elicit accurate responses. Issues to consider include sample size, survey timing, methods to maximize response, and analysis of responses.

Surveys have been a consistent aspect of IM programs. Three distinct stages of survey implementation and analysis are highlighted within this protocol: initial survey considerations, survey design, and analysis methods. Initial survey considerations (e.g. sample size and survey timing) and the survey design process should be decided prior to conducting surveys. Once the surveys are completed, analysis methods such as weighting data and calculating VMR can be employed to provide greater confidence in results.

INITIAL SURVEY CONSIDERATIONS

Before conducting IM surveys, practitioners should follow the recommended best practices described below. Careful consideration of sample size and design, survey timing, and the use of incentives and reminders will result in improved data collection processes and higher response rates.

Sample Size And Design

IM travel surveys often report low numbers of biking, walking, and transit trips. To ensure a proper analysis, it is important to calculate adequate sample sizes. In cases where the number of participants is low, consider using a panel survey that compares the same group of people or households in the pre- and post-program surveys. Designating a control group is also important for isolating the effects of external factors such as weather, changes in gas prices, and new transportation infrastructure. To calculate response rates needed for a particular program, it is necessary to establish a level of confidence,⁷ which is typically at 95 percent.⁸ As the target population becomes bigger, a proportionately smaller sample is required. For example, a target population of 10,000 requires a sample size of 385 to have a 95% level of confidence. It is important to obtain sample sizes that are large enough for any changes in biking, walking, and transit use to be statistically significant.

Panel Survey Benefits

- Analysis can be completed for programs that have smaller sample sizes.
- There are opportunities to revisit with participants on a more frequent basis or over a longer time period.
- The panel can be used for a mixture of quantitative and qualitative research, and for continuous research such as tracking attitudes and opinions over time.

Survey Timing

Transportation behavior can vary greatly based on weather, events, infrastructure disruptions, or other factors. Most IM programs conduct pre and post surveys during similar seasons; thus, the timing between these surveys could range between four to six months or up to one year in duration. Short-term behavior changes associated with an IM program can be captured by conducting a post survey immediately following the intervention, whereas longer-term behavior changes can be captured by allowing more time between the intervention and the post survey. Conducting a third follow-up survey up to six months or one year following the post survey will help determine if behavior changes are sustained over time. However, increasing the duration between pre and post surveys increases the probability for external factors such as changes in infrastructure or gas prices to influence travel behavior.

Surveys conducted during one specific time of year also have limitations, as is the data may not reveal the actual behavior of respondents because transportation decisions may drastically change during different times of the year. For example, people may reduce the amount they drive in the summer because the weather is nice. However, in the winter, they may actually increase their driving so much that it completely cancels out any VMR figures in the summer, and could even increase their net vehicle miles traveled (VMT) for the entire year. Sampling during only one season does not capture this data.

Best Practices For Survey Timing

- Survey respondents during different times of the year if budget allows.
- Compare results from one season to another and/or one year to another.
- Conduct a third follow-up survey to measure longer-term behavior change.

Increasing Response Rates

High survey response rates increase the likelihood of achieving more defensible and statistically significant results. IM travel behavior surveys typically yield between 5 and 25 percent response rates, depending on the type of survey (e.g. paper, online, or phone) the incentive offered, and other factors such as target-area demographics. If a control group is used, it is important that the initial sample size is adequate for this group to maximize survey returns and provide a meaningful analysis. Achieving high response rates from a control group during the post survey is sometimes challenging, as this group typically does not receive any of the program offerings, thereby limiting motivation to respond. Thus, incentives and a well-designed survey are important for increasing response rates.

The following tips will help practitioners increase survey response rates:

- Determine adequate sample sizes for both target and control groups at the outset of a program. A number of online calculators are available to assist with this task.⁹
- Design a survey questionnaire that is easy for the respondent to read, fill out, and return.
- Develop an online version of the survey and promote the URL in mailings as a supplement to mail-based surveys.
- If budgets allow, offer a large incentive for the pre survey and an even larger incentive for the post-survey (e.g. a chance to win a \$100 gift card for the pre survey and a \$200 gift card for the post survey).
- Send out a series of reminder postcards (target and control groups) that includes a URL to the online survey. If you have a designated control group, it is imperative to create a separate online control group survey and URL for tracking those responses separately.
- Use social media, calling, and email reminders to promote the survey for participant-based programs.
- Use a postage-paid reply device.

SURVEY DESIGN

Good survey design provides a more robust data collection process. For example, incorporating standardized survey questions and five-point scale attitudinal questions into IM surveys yields more compelling data. This section includes a series of recommendations and tips for practitioners to use when designing IM surveys.

Standardized Survey Questions

Using standardized survey questions to evaluate IM programs allows results from multiple programs to be aggregated and analyzed over time. Although survey questions can evolve over time, core questions should remain the same. This allows for larger comparisons between pre and post groups that will increase the statistical power of the analysis.

Measuring Mode Shift

Mode shift is a key metric for IM program evaluation and is measured through trip diary questionnaires that are conducted by mail, phone, or online. For IM programs that measure all types of trips (e.g. both commute and discretionary), one-day and multi-day trip diaries are common instruments used to measure mode shift. Other methods can also be employed, such as asking the frequency of trips by different modes over the course of a week or a month. While there are some advantages to this method (e.g. offering a longer time span for reporting travel behavior), accurate recall among survey respondents may affect the quality of the data and reduce response rates. It is also more difficult to collect detailed information about trip purposes (e.g. work, recreation, errands) using these methods because of too much memory recall by respondents.

Benefits Of Single And Multi-day Trip Diaries

Benefits of a one-day trip diary survey:

- Low recall burden
- Achieves high response rates
- A shorter survey allows space for attitudinal questions
- Easier to capture trip purposes for all travel modes

Benefits of a multi-day trip diary survey:

- Ability to capture travel patterns over a longer time period
- Less chance of recording "non-typical" travel days
- Good for commute focused programs (trip purposes not needed)

Survey Design Best Practices

Survey design best practices can be applied to the scale and nature of IM programs. Example survey questionnaire designs from various IM programs are included in Appendix A.

- Add a trip distance category for all trips listed in the trip diary to measure VMR and increase statistical power.
- Break out the carpooling mode into "carpooling with adults" and "carpooling with children" to recognize adult carpool trips versus trips that involve escorting (e.g. picking up and dropping off) children.
- Add five-point scale questions to measure attitudes about transportation options across different sub groups. For example, ask the question "How easy or difficult is it for you to meet your transportation needs by walking, biking, using transit, and/or carpooling? [Very Easy, Easy, Neutral, Difficult, Very Difficult].
- Ask the question "Which of the following transportation options are available within one mile from your residence?" and list available options in the target area as answer choices.¹⁰
- Supplement existing demographic questions by asking for age, income, ethnicity, and zip code. Include the answer choice "prefer not to say" to allow participants to opt out of answering the question and also make the demographic questions optional. Requiring personal information may hinder response rates.
- Design surveys that allow for easy quality control measures such as identifying duplicate responses, outliers, and errors in the data sets and checking to make sure trips are logical.

Retrospective Survey Questions

Retrospective survey questions can be used for shorter-term campaigns such as commuter challenges for which a pre survey is not necessary. Retrospective questions can also be asked in post-program IM surveys to gauge increases or decreases in transportation habits over the course of the program. A list of sample retrospective questions for use in participant-based surveys is included in Appendix C.

Analysis Methods

Once pre and post surveys are conducted, practitioners can further analyze the data to provide greater insights for travel behavior changes and also to produce more defensible results. This section explores techniques practitioners can use to weight data and measure VMR.

Weighting Data

Weighting is an important concept for using data from the sample to reveal insights about a larger population and also to compensate for sampling errors. However, it is important to note that weighting does not improve the quality of the sample. Weighting can be an effective tool for improving analytical accuracy, especially when using survey results to model the travel behavior of a larger area such as a census tract.¹¹

If IM survey data is being used to study or extrapolate trip and VMR behavior across an entire region or city, certain weights can be attained for the following characteristics: ethnicity, zip code, median household income, and other demographic variables. The primary purpose of weighting is to acquire and analyze a sample that is representative of the larger population as a whole, which is accomplished by finding the percentage of each subgroup within populations and applying that coefficient to the samples. For example, if the study wishes to weight the samples on ethnicity, the percentage of each ethnicity in relation to the larger population will be used as the weighting coefficient.¹²

Measuring VMR

IM programs requiring VMR calculations should be measured using a defensible method.¹³ The main goal of reporting on VMR is to find out if participants changed their amount of VMT as a result of being exposed to the program intervention. To accomplish this goal, trip distances should be collected from survey respondents (via an open-ended response or drop down menu) to provide a continuous measure to compare pre to post survey groups and target to control groups. Recording trip distances in both the pre and post surveys is important for measuring the true impact of the program. This method also allows analysts to look for outliers (e.g. extreme values) within data sets. Once a statistically significant change is determined by comparing the average miles driven for each group of survey respondents,¹⁴ practitioners can confidently say the average driving distance is sound.¹⁵

Recording trip distances also provides greater insights for reporting on categorical variables like geography, ethnicity, and household size, without worrying about weakening the data set.¹⁶ For example, an analysis of ethnicity could reveal insights like, "Latinos drive more miles than other ethnicities by a statistically significant margin." Adding more categorical variables also has the potential to reveal whether specific subgroups of survey participants are more likely than other groups to drive more or drive less. This allows for the potential weighting of samples by comparing the categorical variables to another dataset using census data.¹⁷

VMT Measurement Method

1. Collect trip distances for each drive-alone trip in travel diaries.
2. Multiply the number of drive-alone trips for each participant by the average distances determined by the survey participant.
3. Sum all miles traveled for each participant to obtain total VMT for each group (pre and post or experimental and control).
4. Subtract the VMT of the second group (post or target group) from the first group (pre or control group).
5. If the resulting VMT figure is positive, then the program successfully impacted travel behavior (e.g. yielded a decrease in VMT).

A successful program in Milwaukie, Oregon offered a series of events for seniors, including this nature walk along the Willamette River



PROTOCOL # 4. PASSIVE TRACKING

The Takeaway: Mobile apps and wearables are making it easier for individuals to track their own travel behavior (usually for health purposes). The TDM industry will soon be able to take advantage of these and similar tools to evaluate behavior change in the context of TDM efforts. However, the tools available today are only suitable for tracking walking and bicycling activity.

Today, most IM campaigns rely on some form of self-reported survey data to determine if any mode shift has happened. While surveys are accepted as a standard, they also present challenges. They can be expensive and time-consuming to administer. Perhaps more importantly, there are opportunities for bias and error in self-reported survey data. Survey respondents may not be a representative sample of the target area; they may report inaccurate data on purpose (because they want to present what they believe is the “right answer” or because they want an incentive associated with a certain response) or by accident (e.g. forgetting individual trips). Relying on data from one day or week may lead to the influence of external factors (e.g. road construction or unseasonable weather); surveys completed at different times of year may introduce seasonal variables.

For these reasons, passive tracking is considered by many to be the “holy grail” for the TDM industry. Passive tracking involves a program participant granting access to data that is automatically collected by a smart phone app (e.g. Moves or Endomondo) or a wearable (e.g. a FitBit). This would result in a large volume of highly accurate data being collected over time. The benefits of passive tracking are many. It would eliminate response bias and recall errors; it would provide enough data to more accurately detect changes even for modes that have a low baseline mode share; and it would reduce the effect of weather or similar external factors to skew the data. However, because passive tracking requires a smart phone or wearable, and because participants must agree to share data, it brings with it different concerns about sample bias, privacy, and equity.

For this study, we reviewed the current state of passive tracking options that could be used by IM campaigns and other TDM programs. There are a great number of apps and wearables available that track and report bicycling and walking (e.g. Moves, Strava, Endomondo). These are aimed at people who want to track their physical fitness. For certain types of programs, this may be sufficient. For example, a short-term bicycling-focused challenge can make use of these apps and tools today to good effect.

For programs that include transit, however, no app on the market today is able to accurately sense and report transit trips. And it is hard to imagine any app/wearable being able to differentiate between a drive-alone car trip,

a carpool, or use of a taxi/ridehailing service. RideAmigos is developing an app that reportedly will be able to discern transit trips; however, it will only be able to track home-to-work commute trips, and thus is not immediately useful for most IM programs. Other manufacturers will undoubtedly release similar apps in the near future.

There are other barriers to making use of passive tracking for TDM programs today. Users report that any app that runs constantly in the background drains battery life. Some users may also have limited data plans, pay-as-you-go cell phone plans, or limited storage space, and may therefore be reluctant to participate. Some people will not opt in because of privacy concerns, and others will be ineligible to participate because they do not own a smart phone or wearable. In the past, communities of color and lower-income groups have been left out of programs that rely on smart phones, but over time that is changing; it is now primarily older adults who are left out based on technology. And even for those who have a smart phone, it can be difficult to talk someone into downloading an app, setting up an account, and going through the steps needed to allow access.

The project Advisory Committee concluded that all of these current barriers are still too great for passive tracking to replace surveys for IM programs today. There may be potential to use passive tracking as a supplement to surveys (though that presents challenges for integrating two data sets that are extremely different). For research programs that can offer sizeable incentives, it may be possible to persuade more people to download and use an app, at least for a fixed period of time.

However, passive tracking has enormous potential, and TDM practitioners should be monitoring this sector for improvements that will, at some point, result in a product that can be used to great effect in TDM programs.

Characteristics Of An Ideal Passive Tracking App

For passive tracking to be truly useful to IM practitioners, it would need to have the following characteristics:

- Automatically detect walking, bicycling, transit, and driving trips accurately
- Run in the background to record all trips over a given time
- Require minimum battery drain and data use
- Require no actions from the user other than allowing access
- Provide data in a format that is easy for program administrators to use and understand

There is further potential for an app that becomes the very platform for delivering the program. One can imagine the app not only sensing travel behavior, but allowing program participants to set goals, receive feedback, earn incentives, and receive customized messages—and being integrated into travel information or transit payment apps. The rapid pace of technological change means that such an “all inclusive” app may be possible within five years.

PROTOCOL # 5. NEW METRICS

The Takeaway: IM programs have larger impacts than just vehicle miles reduced (VMR). Tracking and reporting on a wider range of metrics can tell a deeper and more nuanced story about the impacts of programs. This section proposes a range of possible metrics related to travel behavior, community, health and safety, and environmental impacts.

There is a demand for measures of success for IM programs that transcend just mode shift and VMR. An expanded set of possible metrics can paint a broader and more complete picture of the impacts of a given program. Some of these alternative measures are still in the realm of travel behavior (but go beyond VMR), while others are related to health, community, and environmental benefits. Developing new metrics will ultimately help practitioners better collaborate with health researchers and academics from other disciplines. Furthermore, measuring and reporting on new and compelling outcomes from IM programs can potentially unlock new types of partnerships and funding opportunities. New metrics are broken down into the four categories listed below:

- Travel behavior metrics
- Community-oriented metrics
- Health metrics
- VMR-related environmental metrics

TRAVEL-BEHAVIOR METRICS

IM programs have the potential to report travel behavior metrics beyond mode shift. The following section describes different travel behavior metrics that can be measured from IM surveys. Furthermore, supplemental data related to the project such as transit, bicycle, and pedestrian counts can be used to validate mode shift results.

Measuring Shared-use Mobility Services

Shared-use mobility services (e.g. carshare, bikeshare) and Transportation Networking Companies (TNCs) like Uber and Lyft are becoming more prevalent throughout cities.¹⁸ Use of these services is increasing; therefore, it is important to think about how to best measure and report on the potential uptake of these services when evaluating IM programs. For example, from an environmental perspective, riding in an Uber is more like a drive-alone trip than a carpooling trip, because a TNC car takes up the same amount of space and produces the same amount of emissions as a single-occupant vehicle. However, a person's decision making process is also affected when a cost per trip is involved, and TNC use has been correlated with reduced car ownership and an improved attitude toward multimodal transportation.¹⁹

More research is needed to learn about the overall impacts of TNCs on travel behavior. For the time being, however, beginning to track these types of “specialized” transportation services will allow for analysis and tracking trends. Simple questions related to these services can easily be asked in both pre and post surveys to track potential increases in TNC and bikeshare trips during an IM program: “How many trips in the past month did you take using bikeshare services? How many trips in the past month did you take using for-hire transportation services (e.g. Taxi, Uber, Lyft)?”

Reduction Of Vehicles In The Household

It is common for IM surveys to ask about the number of vehicles that are owned or leased in the household; however, it is difficult to directly attribute shedding a car as an outcome of an IM program. For example, a household could have sold or given away a car simply because it was old, broken, or needed by a family member or friend outside of the household. A better way to measure car shedding is to ask the question “Did your household sell, donate, or give away one or more motor vehicles as a result of participating in the program?” If respondents answer “yes,” then ask a follow up question to explore the reason(s) why they decided to shed a car. Answer choices might include the availability of car or bike sharing services, purchase of a bus pass, costs involved with owning and maintaining a car, or transportation options information received from the IM program.

Measuring Bicycle Access/Ownership

Asking participants if they have access to a bicycle helps determine potential equity concerns in the target area and can also be used to look more closely at the transportation choices of a sub group within the data set. For example, one might look at travel patterns and the perceived level of confidence in making bike trips for participants who indicated they do have access to a bike. To attribute new bicycle ownership/access directly to a program outcome, practitioners can ask a follow up question: “Did you buy, rent, borrow a bicycle, and/or join bike share as a result of participating in the program?”

Number of New Users of Sustainable Travel Modes

Trying out a new mode of transportation may be the hardest step IM programs ask people to take. Yet most evaluation approaches only measure sustained or habitual use. Measuring whether the program led people to try out a new mode of transportation can reveal whether the program is overcoming major barriers. There are two ways to ask this question. The first lumps three sustainable modes together: “Did you try out a new way of getting to a destination by bicycling, carpooling, or using transit due to the program?” If there is an interest in getting more detailed information about which particular modes were used, the following question can be asked: “Since signing up for the program, which of the following transportation options did you try out?” [Bicycling, Transit, Carpooling, Don’t know].

Using Bike, Pedestrian, and Transit Counts

There is potential for travel survey data from IM programs to be analyzed in conjunction with bike, pedestrian,²⁰ and transit count data²¹ collected from within the target area and during the duration of the program. Although it is challenging to attribute these counts directly to program outcomes because they typically capture travel data from both participants and non-participants, examining external data can provide insights into possible trends associated with IM mode shift results. For example, if an IM program shows a 4% increase in transit and the local bus service in the target area showed a net increase in ridership during the program time span, those findings further validate the mode share increase for bus trips (i.e. the trend is going in the right direction).

Use Of A Distributed Incentive

Some IM programs offer incentives such as free bus passes, transit fare cards,²² and bikeshare memberships.²³ The number of participants who use these incentives should be tracked as an evaluation outcome. If incentives are not directly offered via an IM program, it is important to measure the number of sign-ups/memberships of those services that are available in the target area. For example, if carshare services are available, practitioners should ask the question: "Did you join or become a member of any carshare services (e.g. Zipcar, Car2go) as a result of participating in the program?"

Community-oriented Metrics

IM programs that utilize neighborhood-based outreach strategies such as home deliveries of materials and custom events should attempt to measure potential community- and business-related impacts. The following section offers a series of evaluation tools and questions that may help demonstrate increased local trip making and positive community outcomes from IM programs.

Increased Local Trip Making

IM programs often encourage participants to make multimodal trips within their own neighborhoods. To measure this potential outcome, local business coupons can be tracked if they are offered as part of the IM campaign. Practitioners can also ask survey questions to program participants to confirm that local trip making occurred in the target area. Example questions follow:

"To what extent did the program encourage you to visit nearby shops, restaurants, or parks? [Encouraged me a lot, Encouraged me somewhat, Did not encourage me much, Did not encourage me at all]"

A series of questions can also be asked using a ranking system:

Please rank the following statements with 5 being strongly agree and 1 being strongly disagree (3 is neutral):

- The program encouraged me to walk, bike, take transit, and/or carpool to visit nearby parks.
- The program encouraged me to walk, bike, take transit and/or carpool to visit nearby businesses.
- Because of the program, I am more knowledgeable about nearby services and amenities that are within walking and biking distance from my residence.

Community Cohesion And Connectedness

Community cohesion and connectedness metrics can be measured qualitatively through stakeholder discussion groups or via participant-based focus groups held before and after an IM program to gauge potential increases in community cohesion and connectedness due to the program. The following list of topics can be explored and ranked by participants through these groups:

- Sense of place/community
- Neighborhood safety
- Communications and interactions with neighbors (e.g. more people out on the street biking and walking)
- Involvement in community activities (e.g. neighborhood watch, neighborhood association meetings, National Night Out)
- Participation in Safe Routes To School programs
- Support for local businesses
- Participation in Open Streets Events

Health Benefits

Health outcomes are generally measured in increased minutes of physical activity performed. In the context of an IM program, health outcomes can either be measured directly (by asking survey questions related to levels of physical activity) or calculated by inputting mode shift data into the Integrated Transport Health Impact Model (ITHIM) tool. Various IM programs in the US have incorporated questions about changes in physical activity in pre-and post-program surveys.²⁴

Example physical activity questions for pre- and post-program IM surveys include the following questions. Analysis involves comparing the responses in minutes and miles²⁵ between pre and post surveys.

"Thinking about last week, approximately how many minutes did you spend walking for any of your trips, including walking to public transportation? Approximately how many minutes did you spend bicycling for any of your trips?"

“Thinking about last week, approximately how many miles did you walk for any of your trips, including walking to public transportation? Approximately how many miles did you ride a bike for any of your trips?”

ITHIM—developed by researchers at Cambridge University and increasingly used in the US—is a more sophisticated tool that can calculate health benefits of an IM program using mode shift results and existing county-wide health data. The changes in transportation behavior (measured from area-wide travel diary surveys) can be used as inputs for ITHIM. The tool calculates important health indicators such as increased minutes and distances for walking and bicycling trips, as well as monetary savings resulting from decreased burden of disease.²⁶ ITHIM was used to measure health impacts of six Drive Less Save More IM programs in Oregon and has potential to be replicated in other states.

ITHIM CASE STUDY

Mode shift results from the “Drive Less Save More: SouthTown” Individualized Marketing Program in Corvallis, Oregon showed a 3.8 percent reduction in the drive-alone mode share, a 2.2 percent increase in transit, and a 1.4 percent increase in walking (determined by pre- and post-program trip diary surveys). After inputting this data into ITHIM, it was possible to report that the program resulted in a decrease of -2.1 Disability Adjusted Life Years (a 0.44 percent decrease in the burden of disease) among target area residents, which was valued at \$115,300. The model also showed other health-related outcomes for all SouthTown residents in the target area:

- Participants increased their walking by an average of about 1.25 miles per week.
- The average increase in physical activity was about 4 minutes per person per day, nearly 20 percent of the recommended 150 minutes per week.

VMR-related Metrics

Additional metrics such as environmental benefits²⁷ and cost savings can be easily calculated for programs that measure VMR. Once a VMR number is established, the following data sets can be used as multipliers to convert VMR into cost savings and environmental benefits.²⁸ It is important to note the conversion metrics are based on aggregated, national data sets and results should be interpreted with caution, especially if used in the context of smaller, neighborhood-based IM programs.

Table 1: Automobile operating and ownership costs per VMT



Ownership Costs

Auto type	Cost per VMT/ savings per VMR
Sedan	\$.57
SUV	\$.68



Costs of Crashes and Congestion

Table 2: Costs of crashes and congestion per VMT (2008 values)

Study area size	Cost of crashes and congestion per VMT
Very Large city (over 3 million)	\$.25
Large City (1-3 million)	\$.28
Medium City (500,000-1 million)	\$.32
Small City (500,000)	\$.41
All cities	\$.28

Source: Crashes vs. Congestion - What's the Cost to Society? 2008 AAA report, Figure ES.3: <http://newsroom.aaa.com/wp-content/uploads/2011/10/200835920140.CrashesVsCongestionExecutiveSummary2.28.08.pdf>



Emissions Cost Savings

Table 3: Emissions cost savings

Total Hydrocarbons	Cost/value per VMT
Fine Particular Matter (PM2.5+PM10)	\$.02
Nitrous Oxides (NOx)	\$.01
Carbon Dioxide (CO2)	\$.02
Total emissions savings	\$.05

Source: http://www.nhtsa.gov/staticfiles/rulemaking/pdf/cafe/FRIA_2017-2025.pdf Table VIII-16

PROTOCOL # 6. MAX-SUMO FRAMEWORK

The Takeaway: MAX-SUMO is a TDM evaluation tool used primarily in Sweden. It provides a rigorous and logical method that can be used or adapted for North American TDM programs. It is unlikely that MAX-SUMO or a similar tool will be adopted as the top-down standard in North America. Nevertheless, individual practitioners and regions may find many aspects of the tool useful.

MAX-SUMO is a comprehensive evaluation framework primarily used in Sweden. It is intended to be flexible enough to use on any type of TDM (also called "mobility management") project. While Sweden came to TDM relatively late, rigorous evaluation results have led to it becoming mainstream and integrated into city planning. The tool was adapted for use throughout the European Union, but it has not become the universal evaluation tool that promoters had envisioned.

MAX-SUMO walks practitioners through a series of potential monitoring 'levels,' each of which assesses a unique aspect of the project. It allows practitioners to evaluate the campaign activities, changes in participant awareness, attitudes, satisfaction, and changes in behavior from transportation behavior change programs. The levels fall into three groups:

- Assessment of the service provided (that is, assessment of the program quality)
- Assessment of travel mode(s) offered (that is, assessment of the uptake and assessment of the quality of travel options that people tried)
- Assessment of program impact (changes in attitudes/behaviors and system impacts)

Users are free to select which levels are appropriate for a given program, considering the goals and constraints, but it should be noted that all levels can provide valuable feedback about the program's input. A detailed summary of the benefits of each level of assessment is provided in Appendix F.

A monitoring and evaluation plan created with the MAX-SUMO method will be thorough and defensible. Practitioners can benefit from testing the concepts from MAX-SUMO even if they adapt them for a local context. Like all evaluation tools, MAX-SUMO has minimal value if it is first used at the end of a project. Instead, the monitoring and evaluation plan should be developed at the outset. This allows for the right data to be collected throughout the project.

One unique aspect of MAX-SUMO is that practitioners are encouraged to set targets for each level of assessment. For example, for Level B, the target may be that 45 percent of the target area residents are aware of the project. The process of setting targets is not primarily an exercise for external accountability (such as to funders). Rather, it is intended to stimulate practitioners to think hard about the program, to compare it to other programs and their own experience, and to encourage internal accountability to the vision of success.

In the US context, there is no single, standardized evaluation framework, so using or adapting MAX-SUMO would not (in the near term) bring the comparability and transparency advantages. Nevertheless, because MAX-SUMO is particularly robust and tested, it is worth considering for US TDM program evaluation, particularly for larger and more comprehensive efforts.

There are a number of challenges associated with applying MAX-SUMO as it is today in the US context, including:

- The idea of assessing satisfaction (Level G) with the target mode(s) may be challenging for the US, where TDM practitioners seldom have any input into the quality of transportation systems and services. Swedish evaluation experts admit that this level may not be useful for most programs, but it still can provide valuable context for the program results. See Appendix A for a discussion about assessing this level.
- Some programs in the US only survey participants. This makes it impossible to measure some of the MAX-SUMO indicators, such as awareness of the program among potential participants (Level B).
- Some of the levels require both a post-program survey during or immediately after the program, and a longer-term follow up survey. Many US programs do not have the budget to allow for this much evaluation. However, Swedish experts recommend longer-term follow up surveys where possible, potentially to a smaller subset of participants.
- Many US programs do not have sufficient budget to prioritize surveying at this level. This is likely to be the major barrier to using MAX-SUMO in the US.

The primary challenges to using this framework in the US have to do with the fragmented governance of TDM (where TDM campaigns are often carried out by organizations that don't run transportation service or build transportation infrastructure), and the cost of surveying.

In the US, there are two ways that MAX-SUMO or a tool like it could become an industry standard. One is for the USDOT to establish it as a requirement or strong recommendation. For example, such a tool could become a requirement for CMAQ reporting, or be used as a scoring factor for government grant applications. Alternately, a state or metropolitan planning organization could adapt it as a statewide or region-wide reporting standard (and presumably set aside the needed funding for the evaluation surveys).

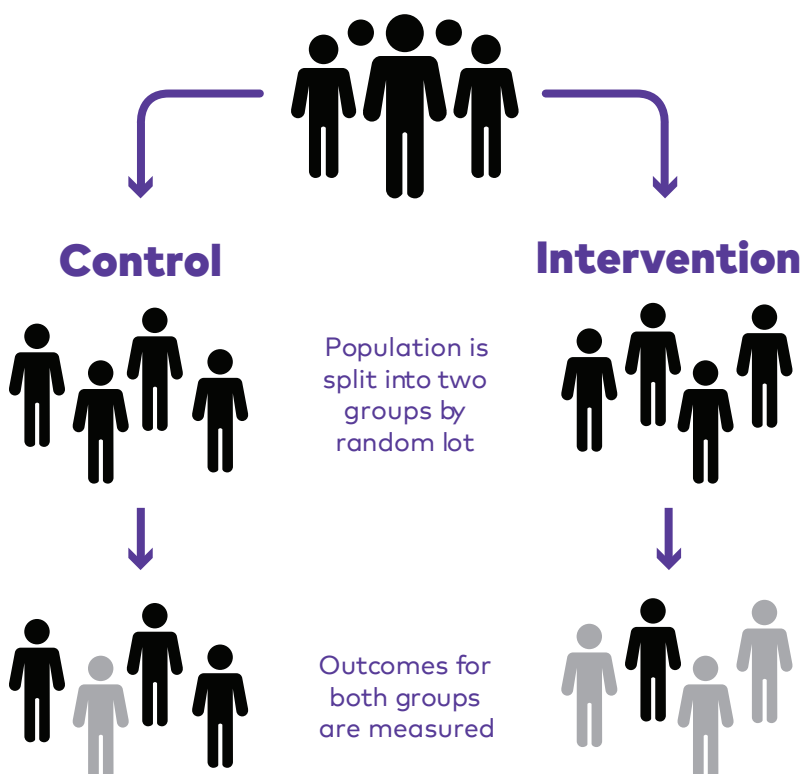
PROTOCOL #7. A/B TESTING

The Takeaway: Using randomized controlled trials to “A/B test” elements of IM campaigns will provide insight into which aspects of programs are most effective. This tool is best used to optimize current programs rather than to evaluate the larger impact of programs on the target audience.

This project focused primarily on evaluating the impacts of overall IM programs. A/B testing offers a different opportunity: evaluating the effectiveness of specific program elements. This allows practitioners to continually improve programs, and to have confidence that they are becoming more cost-effective over time.

The concept behind testing is simple: Try two different options and observe which one is more effective. The target audience may be the entire target area (potential participants), or it may be current participants. For example, perhaps practitioners will want to test whether a postage-paid mail-back order form is worth the additional cost compared to directing people to an online order form. In this case, you would send a postage-paid order form to half of the target group, while the other half would receive a similar mailing that directs people to an online order form. At the end of the program, you would observe what percentage of each group completed an order form, compare that to the cost of the mailing, and decide which option was more cost-effective.

Figure 4: A/B Testing Diagram



To be confident in the results of your test, you must follow a few important principles.

1. Firstly, and most importantly, the groups must be randomly assigned. This means true randomization, not a shortcut; for example, using even/odd addresses to assign groups is not randomization. Each of these groups must receive only the treatment to which they have been assigned. Thus, it is typically not practical to test social media outreach against direct mail because it is profoundly challenging to send a mailer (and only a mailer) to group A and a Facebook ad (and only a Facebook ad) to group B. There are numerous randomization tools online,²⁹ or a random number generator function in Excel can be used to randomly assign numbers, followed by sorting by number and division into two groups.
2. The groups must be large enough to return results that are statistically significant. The exact number needed in each group depends on the size of the difference you expect to see; a smaller difference requires a larger group. (So, if you expect to see a difference of 1 percentage point, you might need a sample size of approximately 31,000 for each group, while if you expect a difference of 10 percentage points, a sample size of 300 is sufficient.) There are numerous online calculators available to determine sample size and statistical significance.³⁰
3. In most cases, you will want to isolate the variable being tested. Without this, you cannot be certain which variable led to any changes you observe in the results.³¹ In the example given above, for example, the two mailings should look alike and be sent in the same month. If mailing A (postage-paid) was a full-color folded mailer that arrived in March, while mailing B (directing people to an online order form) was a black-and-white postcard that arrived in June, it would be difficult to say which variable was responsible for any difference that was observed in the outcome (the mail-back/web medium, the format of the mailing, the use of color, and/or the season/month).
4. The outcome being measured should be easy to observe and very directly tied to the action you took. The results of testing two messages related to the benefits of bicycling more often will be difficult to evaluate with any confidence, because it is difficult to assess mode shift changes in the short term. In the context of IM campaigns, this means that tests related to increasing program participation or completing surveys will be more appropriate (because the results are easy to observe) than tests related to the larger goal of reducing drive-alone trips and increasing the use of active and shared modes.

It does not take particularly advanced skills to set up and evaluate a simple A/B test, but there are some details that you'll want to get right. For this reason, if you do not have experience with experimental design and basic statistics, you may want to find an advisor for your first few tests. Consider contacting your local university, or find out if your agency has a statistician in a different department.

There are many aspects of IM campaigns that could benefit from A/B testing, including increasing program registration, survey completion rates, and efficiency (e.g. reducing the number of mailings or the cost and type of incentives used). These could be tested with different messages, different media (e.g. phone call vs. direct mail), different formats (e.g. postcard vs. mailer), timing of contact, or incentive options (e.g. smaller incentives for all vs. a larger drawing for a single recipient of the incentive).

A/B testing offers a unique tool for IM practitioners to build a knowledge base about what works, and also to optimize programs for efficiency and effectiveness. We were not able to find tests along these lines that have been done to date, but the TDM industry stands to benefit greatly from this idea. The most good will come if practitioners who do try out testing are able to share their findings with other practitioners. While there is no IM network at present, the Center for Urban Transportation Research's TDM listserv is an efficient way to share information, and conferences and blog posts can also be of use to many types of program implementers.

Bicycle workshops and rides to Saturday Markets attract families and are beneficial for highlighting low-stress bike routes throughout neighborhoods



FIVE RECOMMENDATIONS FOR BETTER BEHAVIOR CHANGE PROGRAMS

New technology, in the form of enterprise-level software, transportation apps, and emerging on-demand and shared mobility companies, offers the potential for better behavior change programs.

But to be effective, TDM professionals must be committed to thorough and transparent evaluation of programs. Further, the results of evaluations must be shared more readily and widely to benefit the field as a whole. This report aims to be a catalyst for new and different ways of evaluating the success of TDM programs. We invite you to use, share, change, critique, and build on the protocols we have developed. We hope thereby to make a contribution to stronger, more defensible, more cost-effective ways to change transportation behavior.

Beyond the specific guidance to practitioners outlined above, the project team identified five big ideas related to IM campaigns and TDM.

1. CITIES BENEFIT BY GOING BIG ON BEHAVIOR-CHANGE CAMPAIGNS

Most IM campaigns are relatively small. A typical first-year program might serve 5,000 households. This is primarily because programs are grant funded, and a typical grant award amount might be less than \$200,000; working backwards, that leads to target areas of 5,000 to 8,000 households. There are disadvantages to a program of this size, however. The time needed to set up a program is largely fixed, and printing costs decrease per unit printed as print runs are larger; both of these factors mean that a smaller target area is more expensive per household. From an evaluation perspective, most of the costs are related to evaluation design and analysis, which do not scale down for smaller projects. Smaller target areas are less likely to make measurable system impacts (such as increasing transit ridership to a degree that can be measured by boarding counts). Perhaps most importantly, larger data sets allow for much more robust analysis, especially when it comes to looking at the behavior of sub-groups within the target area.³² And ultimately, TDM programs don't make enough of a difference to matter if they aren't reaching enough people. For all of these reasons, we recommend larger target areas of at least 10,000 households, if not more.

2. ALIGN BEHAVIOR CHANGE SCIENCE AND TRANSPORTATION PRACTICE

TDM practitioners need to better align with professionals in the fields of psychology, behavioral economics, health research, and other behavioral science disciplines to better understand how behavior change occurs in individuals and how to make it a lasting effect. The protocols developed

in this report provide some context for using behavior change theories; however, more research is needed to pilot new concepts that will lead to more innovative and effective program designs and evaluation strategies.

Many cities and agencies are beginning to apply a behavioral lens to their programs, and often are able to provide “nudges” that improve outcomes at little to no cost. The TDM field stands to benefit greatly by learning about the state of the research and beginning to apply it to programs and projects.

3. ALIGN ACADEMIA AND PRACTITIONER INTERESTS

As we show, IM project evaluation suffers from a lack of rigor around sampling and analysis, which is often due to budget constraints. Academic researchers need to become more involved with IM evaluation efforts by aligning their research needs with actual projects in the field. This would offer an excellent opportunity for graduate students to take on survey design work and analyses, which helps improve transparency and defensibility of project results. IM practitioners should pilot this concept by advocating for additional evaluation funding (beyond an often-used “rule of thumb” of 10 percent) in grant applications and through developing partnerships with academics, statisticians, and local research institutions.

4. LOOK BEYOND FEDERAL AIR QUALITY FUNDS

The federal Congestion Mitigation/Air Quality (CMAQ) funding program has been an important ongoing revenue source for many TDM efforts. In many regions, it is the only funding spent on transportation behavior change programs. As a result, TDM evaluation and practice has been slanted (perhaps understandably) to focus on the metrics that are important to CMAQ: vehicle miles reduced and associated emissions reductions. These are important metrics, but they are not the only ways of understanding and talking about the impact of demand management programs.

An exclusive focus on VMR misses the opportunity to talk about other aspects of behavior change such as trying a new mode, making a mental shift in motivation, shedding a car, or shifting from an occasional to a regular user of travel options. And a narrow focus on emissions reductions misses the chance to assess a program’s impact on health, the local economy, quality of life, and community. By bringing TDM into a larger frame, practitioners can build new partnerships and make a compelling case for investments in TDM, from new funding sources like public health departments.

5. INTEGRATE DEMAND MANAGEMENT INTO LOCAL TRANSPORTATION DEPARTMENTS' VISION, GOALS, AND PROCESSES

In many places, TDM staff have little interaction with the professionals who manage street design, public transit, and parking policy, even though those departments play an enormous role in travel behavior. This puts TDM professionals in a position where they are charged with increasing use of sustainable transportation even as the broader policy environment makes this difficult (i.e. because transit is being cut, roads are being widened, or abundant free parking is provided). Often, the only funding is grant-based, making it difficult to undertake longer-term or more complex projects. To truly transform how people move in cities requires different aspects of the system working in concert: policy, planning, engineering, maintenance, monitoring; parking management, zoning and development code, street design, and growth plans; vehicle, freight, transit, shared mobility and active modes. All of these approaches must be aligned to bring about meaningful change.

Demand management can be an organizing principle that guides how cities evolve. For example, the Move Seattle plan adopted by the Seattle Department of Transportation sets a goal of having 72% of city residents within walking distance of frequent transit. With this high-level goal guiding city departments' work, it is easier to justify spending on behavior change programs, and easier to align them with relevant street and transit projects. Communities that set mode shift goals first, then craft policies, plans, budgets, and work plans to accomplish that end, will be best equipped to meet the future mobility needs of all their residents while reducing drive-alone trips.

Commuters in Seattle, WA



ENDNOTES

1. PTP-Cycle projects in Europe target more than 10,000 households per project area: http://www.polisnetwork.eu/uploads/Modules/PublicDocuments/ptp-cycle_final-report_march2016.pdf
2. Cohen, Tom. (2009). Evaluating personal travel planning: If it is prohibitively expensive to get a robust answer then what should we do? *Transport Policy* 16. October: 344–347
3. Using a Logic Model. 2016. <http://toolkit.pellinstitute.org/evaluation-guide/plan-budget/using-a-logic-model/>
4. Choose an Evaluation Design. 2016. <http://toolkit.pellinstitute.org/evaluation-guide/plan-budget/choose-an-evaluation-design/>
5. W.K. Kellogg Foundation Logic Model Development Guide. January 2004. <http://www.smartgivers.org/uploads/logicmodelguidepdf.pdf>.
6. If the only local transit option is bus, then "use transit" may be replaced by the simpler "bus" if desired. If the program is promoting only active transportation, "walk/bike" may be used.
7. Sample size dictates the amount of information collected, which therefore helps determine the precision (e.g. level of confidence) in the sample estimates.
8. An online sample size calculator can be found here: <http://www.surveysystem.com/sscalc.htm>
9. For example: <http://www.surveysystem.com/sscalc.htm>
10. Knowing which transportation options are easily and readily available to participants can provide a more meaningful analysis and interpretation of results.
11. Information about sampling weights and adjustments can be found by visiting the following web links: https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/subject_areas/omnibus_surveys/household_survey/2009/october/html/sampling_weights.html and <http://www.applied-survey-methods.com/weight.html>
12. For a target population of 75 percent white and 25 percent non-white, the sample should contain three white participants for every one non-white participant
13. An example VMR calculation method is included in Appendix D.
14. Running a t-test would compare the averages of the pre and post group and use the standard deviation to say whether the difference in VMT is statistically significant. If the t-test yields a statistically significant result, then the program successfully impacted travel behavior. An Excel t-test example is included in Appendix E.
15. For example, the pre group drove an average of 8.5 miles, and the post group drove an average of 7 miles. If the difference between the groups is statistically significant, then it can be said that a person exposed to the program, on average, will reduce their VMT by 1.5 miles (increase their VMR by 1.5 miles).
16. The best way to balance the structure of a data set is to have a number of continuous, numerical variables such as trip distance combined with a number of categorical variables such as gender and income.
17. For example, if the study wishes to weight the samples on ethnicity, the percentage of each ethnicity in relation to the larger population will be used as the weighting coefficient.
18. A calculator for exploring the benefits of pursuing shared mobility in select US cities can be found here: <http://calculator.sharedusemobilitycenter.org/#/>
19. <https://www.apta.com/resources/reportsandpublications/Documents/APTA-Shared-Mobility.pdf>

20. The St. Paul Minnesota IM programs employed before and after pedestrian and bicycle counts in the target area.
21. The Go Bronzeville IM program in Chicago uses transit counts to complement mode shift data.
22. ORCA cards pay for buses and trains in the Puget Sound area and have been offered as part of Metro's InMotion IM programs.
23. The Go Bronzeville IM program tracks Divvy bikeshare memberships.
24. The Spokane, Washington IM program asked participants how many minutes and miles they walk, bike, or ride the bus in an average week in both pre- and post-program surveys. The St. Paul IM program asked about the prevalence of exercise and health in their travel survey.
25. Walking burns 77.33 calories per mile and bicycling burns 38 calories per mile: <http://golafayette.org/how-are-these-calculated/>
26. Woodcock J, Edwards P, Tonne C, Armstrong BG, Ashiru O, Banister D, et al. 2009. Health and Climate Change 2 Public Health Benefits of Strategies to Reduce Greenhouse-gas Emissions: Urban Land Transport. *Lancet* 374: 1930–1943.
27. The EPA greenhouse gas equivalencies calculator can be found here: <https://www.epa.gov/energy/ghg-equivalencies-calculator-calculations-and-references>
28. CO2 emissions per mile are .95 lbs. (small car) 1.1 lbs. (medium car), and 1.57 lbs. (SUV). An online carbon calculator can be found here: <http://www.carbonify.com/carbon-calculator.htm>
29. <https://www.random.org/lists/>
30. <http://www.surveysystem.com/sscalc.htm>
31. At this point in time, there is no rigorously-tested body of evidence to draw on for IM campaigns, and so it is essential to develop basic knowledge by isolating variables. In the future, when we have more knowledge, there may be situations where we apply multiple principles at the same time in the interest of creating packages of interventions that can be tested against each other to compare impact; in this situation, it is not essential to isolate the variable.
32. For example, Dr. Jennifer Dill developed a meta analysis of multi-year City of Portland Smart Trips programs. Dill, Jennifer and Mohr, Cynthia (2010). Long-Term Evaluation of Individualized Marketing Programs for Travel Demand Management. OTREC final research report, July 2010.

