UNIVERSITY OF MINNESOTA

Nonmotorized Transportation Pilot Program Evaluation Study

Phase 2

CTS 11-13

May 2011

Prepared by:

Center for Transportation Studies

Thomas Götschi
Kevin J. Krizek

Laurie McGinnis
Jan Lucke
Joe Barbeau

In collaboration with:

NuStats – *Austin, Texas*

CENTER FOR TRANSPORTATION STUDIES

Active Communities / Transportation Research Group

U.S. Department of Transportation
Federal Highway Administration

Technical Report Documentation Page

1. Report No. CTS 11-13	2.	3. Recipients Accession No.
4. Title and Subtitle Nonmotorized Transportation Pilot Program Evaluation Study, Phase 2		5. Report Date May 2011 6.
7. Author(s) Thomas Götschi, Kevin J. Krizek, Laurie McGinnis, Jan Lucke, Joe Barbeau		8. Performing Organization Report No.
9. Performing Organization Name and Address Center for Transportation Studies University of Minnesota 200 Transportation and Safety Building 511 Washington Ave. SE Minneapolis, MN 55455		10. Project/Task/Work Unit No. CTS Project # 2007026 11. Contract (C) or Grant (G) No.
12. Sponsoring Organization Name and Address Federal Highway Administration 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code

15. Supplementary Notes
http://www.cts.umn.edu/Publications/ResearchReports

16. Abstract (Limit: 250 words)

The Nonmotorized Transportation Pilot Program (NTPP) is a congressionally mandated program (SAFETEA-LU Section 1807) that, since 2006, has provided roughly $25 million each to four communities—Columbia, Missouri; Marin County, California; Minneapolis area, Minnesota; Sheboygan County, Wisconsin—to spur levels of walking and cycling via a variety of planning measures.

The University of Minnesota Center for Transportation Studies is leading the community-wide population surveys for the Nonmotorized Transportation Pilot Program (NTPP), specifically in phase 2, to measure changes in levels of walking and bicycling as a result of the enhanced conditions for walking and bicycling. To evaluate impacts of the program, two community-wide surveys were conducted before (phase 1: 2006) and after (phase 2: 2010) the pilot program.

This report describes the evaluation efforts based on community-wide population surveys. In contrast to project-specific evaluations, community-wide surveys serve the purpose of representatively assessing community-wide levels of nonmotorized travel behavior, which serve as the foundation for subsequent benefit calculations.

The survey in phase 1 consisted of a short mail-out questionnaire and a computer assisted telephone interview (CATI) among respondents to the short questionnaire. In phase 2 the short questionnaire was integrated in the CATI. The final sample in phase 1 consisted of 1279 complete records and in phase 2 of 1807 complete records.

Statistical analysis focused on evaluating differences between phase 1 and phase 2 in the core variables on nonmotorized travel behavior. The detailed analysis did not reveal any consistent or statistically significant differences between phases 1 and 2. It is important to point out that the inability to detect significant patterns of change is not synonymous to no change occurring. The report discusses some of the factors that make this type of research challenging.

17. Document Analysis/Descriptors Active transportation, Analysis, Before and after studies, Bicycling, Pedestrian, Walking, Infrastructure, Impacts, Performance measurement, Surveys, Travel demand, Nonmotorized transportation		18. Availability Statement No restrictions. Document available from: National Technical Information Services, Alexandria, Virginia 22312	
19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 30	22. Price

Nonmotorized Transportation Pilot Program Evaluation Study, Phase 2

Final Report

Prepared by:

Thomas Götschi, Ph.D
Kevin Krizek, Ph.D

Laurie McGinnis, Director
Jan Lucke, Manager
Joe Barbeau, Program Coordinator

University of Minnesota
Center for Transportation Studies

May 2011

Published by:

Center for Transportation Studies
University of Minnesota
511 Washington Avenue SE, Suite 200
Minneapolis, MN 55455

For more information on this project, please visit:
www.cts.umn.edu/Research/ProjectDetail.html?id=2007026

This report represents the results of research conducted by the authors and does not necessarily represent the views or policies of the University of Minnesota or the Federal Highway Administration.

Acknowledgments

The authors wish to acknowledge those who made this research possible. The study was funded by the Federal Highway Administration (FHWA). The research team would like to thank Gabe Rousseau of FHWA and Ben Rasmussen and William Lyons from Volpe National Transportation Systems Center for their guidance throughout the project. In addition, the research team appreciates the effort of the Nonmotorized Transportation Pilot Program Working Group, which provided key input for the survey design and methodology.

Cover photograph: Matt Miranda

Table of Contents

Executive summary

Background

The Nonmotorized Transportation Pilot Program (NTPP) is a congressionally mandated program (SAFETEA-LU Section 1807) that, since 2006, has provided roughly $25 million each to four communities—Columbia, Missouri; Marin County, California; Minneapolis area, Minnesota; Sheboygan County, Wisconsin—to spur levels of walking and cycling via a variety of planning measures. The central intent of the evaluation component of the program was to measure changes in levels of walking and bicycling as a response to the enhanced conditions for walking and bicycling that were provided.

To evaluate impacts of the program, two community-wide surveys were conducted before (phase 1: 2006) and after (phase 2: 2010) the duration[1] of the NTPP in each of the four pilot communities. Surveys were also administered in Spokane, Washington, which aimed to serve as a control community. The main goal of the evaluation component was to assess increases in the use of nonmotorized travel, in addition to several other aspects potentially affected by the investments. Further, the U.S. Congress requested a valuation of benefits resulting from walking and bicycling, in particular in terms of emission reduction, fuel savings, and health benefits, among others.

This report describes the evaluation efforts based on community-wide population surveys. (Additional evaluations were conducted for individual projects and are described elsewhere.) In contrast to project-specific evaluations, community-wide surveys serve the purpose of representatively assessing community-wide levels of nonmotorized travel behavior, which serve as the foundation for subsequent benefit calculations.

Methodology

The survey in phase 1 consisted of a short mail-out questionnaire and a computer-assisted telephone interview (CATI) with respondents to the short questionnaire. In phase 2, the short questionnaire was integrated in the CATI and efforts were focused on directly recruiting participants for the full survey.

The final sample in phase 1 consisted of 1279 complete records, and in phase 2, 1807 complete records.

[1] Due to extensions to SAFETEA-LU the pilot communities received additional funding and the program was extended beyond the originally set timeline.

The surveys addressed a wide range of transportation-related questions and consisted of 54 variables in total. Distance traveled by trips of particular modes—a crucial variable for benefit calculations—was assessed by asking participants for origin and destination of a single reference trip for each respondent (driving, transit, walking, cycling). To collect equal amounts of information on reference trips for each mode, sampling strategies and hierarchies were applied to assure sufficient records for each, in particular the less frequent modes such as transit and cycling.

Statistical analysis focused on evaluating differences of nonmotorized travel between the core variables measured in phase 1 and phase 2. Sampling weights were applied to adjust for the oversampling of rare modes.

Findings

The detailed analysis did not reveal any consistent or statistically significant differences between phase 1 and phase 2. Possible explanations for these null-findings are discussed in detail in this report, followed by specific lessons learned. These are summarized below. It is important to point out that the inability to detect significant patterns of change is not synonymous to no change occurring. The report discusses some of the factors that make this type of research challenging.

Discussion

The evaluation of a possible impact of the NTPP was difficult for a combination of reasons, some of which are specific to the design of the program and others inherent to analyses of changes in (travel) behavior over time.

Elements of NTPP's design—the short evaluation period, the selection of diverse communities and projects, and the relatively limited resources allocated to data collection and evaluation—have been most challenging to a successful scientific evaluation. This evaluation also revealed that without the availability of routinely collected key data on nonmotorized travel behavior—in particular, traffic counts—intervention evaluations of changes over time (such as this one) pose a scientific challenge too difficult to tackle with the typically available resources for this type of project.

Evaluating interventions that consist mainly of infrastructure projects aiming to target behaviors that are difficult to measure in the first place (such as walking and cycling) is an inherently challenging endeavor. The infrastructure projects require rigid planning, significant allocation of funds, and time to mature to be successful.

Notwithstanding the null-findings presented in this analysis, there are a number of lessons to be learned from this evaluation effort.

Lessons learned

- Impacts of nonmotorized transportation (NMT) interventions are difficult to demonstrate scientifically unless adequate means, resources, and time are allocated. In particular, pre–post evaluations that employ probability-based samples are extremely challenging without the availability of routinely collected data, such as regularly conducted household travel surveys and traffic counts.
- Political and administrative considerations in the planning stage of a pilot program may not always underscore the importance of scientific evaluation. It is therefore suggested that a platform for robust evaluation be considered a critical part of the planning process for such projects from their conception.
- Scientific evaluation of the effectiveness of pilot programs requires adequate resources and evaluation, and it is important that such resources be considered in the planning process from the onset of such a program.
- Scientific evidence on the effectiveness of NMT interventions is growing rapidly. Before engaging in evaluation efforts, it is highly advisable to consider existing evidence to avoid unnecessary replication of efforts.
- The scientific community is advised to invest in the development of data-collection and evaluation methodologies that are practice-friendly, cost-effective, and robust.
- The federal government should seek an active role in advancing and coordinating routine data-collection efforts for NMT to track and evaluate performance of NMT nationwide, as is the case for other modes.

1 Methodology

1.1 General approach

The central goal of this evaluation study was to estimate the impact of the federally funded investments in bicycling and walking in the four pilot communities.

The evaluation of the Nonmotorized Transportation Pilot Program (NTPP) presented in this report is based on two cross-sectional population surveys conducted in 2006 (phase 1) and 2010 (phase 2).

Each survey was conducted in the four pilot communities, as well as in Spokane, which was intended to serve as a single control community.

The survey methodology and general approach have been described in detail in separate reports.

1.2 Sampling methodology

For a description of the sampling procedure, in particular the oversampling and mode-specific sampling hierarchies, see survey methodology report.

1.3 Survey instruments

The survey instruments (questionnaires) used in the surveys can be found in the survey methodology report.

1.4 Statistical analysis

Where possible, phase 1 and phase 2 data were treated equally. Due to some methodological differences in how the survey was administered or how variables were coded between the waves, small deviations occurred. Reanalyzing data from phase 1, however, assures that the analysis between the two surveys is comparable. Therefore, when there are discrepancies between these and earlier reported results (e.g., interim report to congress), this analysis should be used.

Wherever reasonable, results were stratified by community and mode.

Overview of the data

In phase 1, mail-outs of a short questionnaire (self-mailer) yielded a relatively low response rate of 13.5%. Approximately one-third (28%) of participants who mailed back the initial questionnaire afterwards participated in and completed a telephone interview (computer-assisted telephone interview, or CATI).

Owing to the relatively low response rates that were achieved as part of the phase 1 data collection, the short questionnaire was dropped in phase 2 and resources shifted towards increasing the number of completed CATIs, resulting in an almost 50% larger sample in phase 2 using the CATI.

Table 1. Samples obtained for phase 1 and phase 2 surveys

	Phase 1	**Phase 2**
Number of complete records	4,457 self-mailer short questionnaires 1,279 CATI	1,807 CATI (no self-mailer was sent out in phase 2)
Number of variables collected	54	54
Response rate	Self-mailer 13.5% (returned SM/mail-outs) CATI 28% (complete CATI/returned SM)	12% (complete CATI/mail-outs)

For more detailed sample statistics, see the survey methodology report.

Weighting

In efforts to systematically maximize data collected from participants who engage in walking or cycling, the analysis employed a targeted oversampling strategy (see survey methodology report). Oversampling efforts therefore require proper assignment of weights to each record. The purpose of weighting is to assure that each record in the sample receives the equivalent weight it would have received if the sample were drawn randomly. Details of the weighting process are more fully described in the appendix.

Exclusion criteria

In phase 1, 964 records from self-mail questionnaires and 79 CATI collected after December 1, 2006, were excluded from the analysis to avoid overly strong influence of winter climate on survey outcomes. All 1,842 samples collected in phase 2 (through November 2010) were included in the analysis.

Analysis of changes between phase 1 and phase 2

Data were analyzed in STATA statistical software package version 11. Data were treated as two-stage survey data (using the *svy* command family with census tracts and individuals as primary and secondary sampling units), stratified by survey and community, with sampling weights and adjustment for finite population.

Difference between surveys (i.e., the occurrence of changes over time, between phase 1 and phase 2) was tested using a Wald test for the linear combination of Survey2 – Survey1 = 0 (testing whether the difference between surveys is significantly different from 0 at a confidence level of 95%). Positive changes can therefore be interpreted as increases, negative changes as decreases.

1.5 Types of variables

The analysis and the accompanying report, generally, refer to four types of variables or outcomes.

- Primary output variables are direct compilations of questions from the survey and revolve around patterns of use for the modes (amount of the sample, frequency, length for different elements of travel behavior).
- Secondary output variables are largely direct compilations from key questions but may be secondary in purpose or require an intermediate step to arrive at a result.
- Calculated variables are those that require processing two or more primary output variables to arrive at a result.
- Benefit calculations have not been conducted, but proposed calculation approaches are briefly described in the appendices.

Detailed descriptions of these variables can be found in Appendix sections 5.3-5.6.

2 Results

2.1 Description

To assess the impacts of the NTPP on aspects of walking and cycling, the final analysis examines a selection of 30 most relevant variables, including:

- *How many days per week do you walk for at least 10 minutes at a time?*
- *How many days per week do you bicycle for at least 10 minutes at a time?*
- *How did you get from your starting point to the transit stop?*
- *And when you got off the bus, how did you reach your final destination?*
- *Minutes biked per week (bikers only)*
- *Minutes biked per week (all)*
- *Miles traveled per year by: Bike*
- *Miles traveled per year by: Car*
- *Miles traveled per year by: Public transportation*
- *Miles traveled per year by: Walking*
- *Percent of all trips, by mode*
- *How many total places did you visit yesterday?*
- *Of the places you visited yesterday, how many did you visit by: Bicycle?*
- *Of the places you visited yesterday, how many did you visit by: Public transportation?*
- *Of the places you visited yesterday, how many did you visit by: Private vehicle?*
- *Of the places you visited yesterday, how many did you visit by: Ride share?*
- *Of the places you visited yesterday, how many did you visit by: Walking?*
- *...most recent [use of]: Private vehicle*
- *Days since last used: Private vehicle*
- *...most recent [use of]: Transit*
- *Days since last used: Transit*
- *...most recent [use of]: Bicycle*
- *Days since last used: Bicycle*
- *...most recent [use of]: Walk to a destination*
- *Days since last used: Walk to a destination*
- *...most recent [use of]: Walk for recreation*
- *Days since last used: Walk for recreation*
- *Reference trip in miles*
- *Minutes walked per week*
- *Minutes walked per week (incl. all)*

The analysis involved comparing pre–post measures (phase 1–phase 2) for each of the above variables. The analysis is available here, indicating values and measures of statistical significance for baseline (phase 1), bookend (phase 2), and change. The labeling used in the results tables consists of:

2.2 Description of findings

Overall, the results from the phase 1 – phase 2 analysis suggest inconclusive findings. Several factors likely contribute to the inability to detect consistent and statistically significant impacts of the evaluated intervention. It is important to point out that not detecting an impact is not equivalent to lack of impact; therefore, the findings of the analyzed surveys must be considered inconclusive.

Central issues to consider when interpreting these findings are discussed in the following text. In many cases, a single one of these issues could be sufficient to declare the results inconclusive, or "null-findings." For the sake of comprehensiveness, key reasons for why the results could be considered inconclusive are provided below.

Multiple comparisons

Compilations of data from a subset of the survey questions—the primary and secondary output variables (described previously)—provided a wealth of analyses. There are 30 variables for which changes between baseline and bookend surveys have been analyzed. These are done for the four pilot communities combined, each pilot community separately, and in addition, for the control community, Spokane. Some variables, in addition, have been analyzed separately by mode. Overall, the combination of available variables to be analyzed yields more than 400 phase 1 – phase 2 comparisons.

When exploring the results for significance, one needs to keep in mind that, given the large number of comparisons, 10% (approximately 40) are expected to be significant (at the 90% level) *by chance alone*. Alternatively, out of 100 confidence intervals at the 90% percent level, 10 are expected to not include the true value (the true value in this case being a change different from zero).

Out of 421 pre–post comparisons (significance tests), 110 yielded significant changes at the 90% confidence limit. Of those, it is prudent to ignore approximately 75 a priori because they stem from relatively small categories (i.e., ride share, 9), from variables where change is not of interest (total places visited, 4), in categories of non-response (6), or in the variables of "most recent use," which distinguish 5 categories that make it extremely likely for significant changes to occur. (The latter have been aggregated in continuous variables of "days since last use" for each mode to facilitate interpretation.) The analysis detected 35 significant changes between phase 1 and phase 2, which is well within the range of what would be expected by chance alone.

Direction of change

When evaluating whether significant changes could be attributed to the NTPP (from here on referred to as the intervention), a first criterion to apply is whether the observed changes are in the expected direction. Among the 30 analyzed variables, one can distinguish those for which the expected direction of change is obvious (walking, bicycling, driving), some for which it is somewhat ambiguous (e.g., transit), and some variables for which the expected direction of change depends on subcategories (i.e., mode share, trip length, and "most recent use," which distinguishes different time periods in the past).

While significant results in an unexpected direction cannot be declared invalid per se, it would be equally inappropriate to consider significant findings in the expected direction as meaningful in the presence of a comparable number of significant findings in the opposite direction. Based on these considerations, there is no consistent pattern among the observed significant changes between phase 1 and phase 2.

Control community

For changes that may be considered attributable to the intervention in the pilot communities, one would theoretically expect no changes in these same variables in Spokane (it is intended to serve as a control).

Overall, Spokane yields about one-fifth of all significant changes, which is what would be expected if significant findings occur due to chance and not due to the intervention. Therefore the observed results do not indicate any differences between the control community and the pilot communities that could be attributed to the intervention.

Consistency of observed changes

It is important to keep in mind that statistical tests of significance evaluate the difference in the two data sets, but that the existence of a statistical difference does not imply any direct inference on whether the difference is real. Any observed significant changes therefore need to be scrutinized to that regard.

Besides the statistical significance of observed changes, a next criterion to apply when interpreting the findings would be whether observed changes are consistent across the pilot communities, as one would expect.

Magnitude of observed changes

Similarly, it would be important to check whether the observed changes are within the range of what could be expected, given the intervention.

2.3 Discussion and conclusions

Based on the above-described considerations, the phase 1 – phase 2 analysis of the population-based surveys of the NTPP did not reveal any consistent patterns that could be attributed to the intervention. A number of factors could potentially be responsible for the inability to detect significant impacts owing to the intervention. The most important ones to consider are discussed below.

Difficulty detecting an impact employing the population-based survey

It is possible that the intervention had impacts on levels of bicycle and pedestrian use, but for a range of reasons, the impacts could not be detected using the measurement approaches that were employed. To better understand reasons why no change was detected, it is helpful to distinguish between two phenomena:

- Confounding variables or events may have blurred, diluted, or otherwise obscured a "real" impact.
- There are methodological limitations that impede detecting an existing impact.

We briefly comment on each below.

Counterproductive outside factors

External factors could have neutralized the impacts of the intervention. Candidates are any factors that changed between baseline and bookend surveys and could have an impact on the assessed variables, i.e., gas prices, economic conditions, or changes in infrastructure, to name a few. For example, it is widely recognized that the onset of the current economic recession was at the end of 2007 and continued through phase 2. Economic recessions are usually accompanied by lower rates of travel overall, which may be one contributing factor for lack of detectable changes.

Methodological and other challenges to detect existing impacts

Besides a thorough consideration of "causal factors" or "real-world factors," it is equally important to understand and consider possible methodological reasons for the lack of detection of consistent patterns attributable to the intervention. In the following, several important factors are discussed, although it is for the most part not possible to separate their particular relevance for the presented findings.

Effect size

As has been described elsewhere (Krizek, Handy et al. 2009), endeavoring to statistically detect significant changes in nonmotorized travel using probability-based samples is extremely difficult. The smaller the effect, the harder it is to detect statistically. There is limited precedence to judge whether the effects to be expected from the NTPP are "objectively" small.

For example, Portland, OR, as probably the only well-documented example of investments into nonmotorized transportation (NMT) in the United States, has achieved relatively large, five-fold increases in bicycling by investing approximately $60 million over 18 years (City of Portland Office of the Auditor 2007; City of Portland 2008; City of Portland Office of Transportation 2010; Gotschi 2011). Such aggressive gains were largely measured using non-probability-based sampling designs (counts across bridges) and were acquired over a timeframe that is more than four times as long as the duration for the NTPP analysis.

Relative to the Portland, OR case, measuring the community-wide effects in the pilot communities is a tall order: (1) each community, using NTPP funds, will eventually invest roughly a third of the sum that Portland did, (2) the impacts were measured over a period of four years, and (3) the measured impacts were split between bicycling and walking. These combined factors suggest changes are expected to be much smaller. Of most concern among those three factors is the time period for which infrastructure projects have been in place and open to users, given that some projects were not completed by the time of the bookend survey.

Ability to detect a small effect size

How prevalent a particular behavior is in the community is critical to the ability of inferential statistics to measure such. Most statistics that are presented (such as those mentioned above) are based on two assumptions: (1) the sample was random and (2) the responses to the survey have roughly a 50-50 split for most of the questions (for example, 50% will favor a position, 50% will oppose). Aiming to learn about rare events (such as levels of walking and cycling) undermines these assumptions. For example, if the survey wants to know more about women who cycle after sundown, the incident rate is typically extremely low; this suggests that the endeavor will require a larger sample size to reliably detect information about this relatively rare behavior.

Lagged response

Returns from transport infrastructure investments typically follow an S-shaped curve, yielding relatively smaller returns early on but maximum returns only after a certain uptake period. (Levinson, Karamalaputi et al. 2003). It is quite likely the investments in NTPP have not reached their maximum return in terms of nonmotorized travel; in fact, in some of the communities, the interventions were not on-line until early 2010, leaving their effect considerably lesser than what could be expected in the longer run.

Sampling strategy

When collecting data of this nature, there are two main types of sampling frames—probability and nonprobability. Employing probability samples—that is, random samples—is a fairly straightforward task in efforts to generalize results from the sample to the wider population; it is not as straightforward with nonprobability samples.

Surveys of cyclists and pedestrians frequently use non-probability-based samples (like snowball or convenience techniques) because both are easier to administer and have higher likelihood of attracting respondents. Critical issues to be aware of with non-probability-based sampling include: (a) the sample might have a lack of representativeness—and it is often impossible to discern whether or not this is the case; (b) it might be difficult to obtain information about a particular population—for example, an intercept survey fails to learn anything about people not using a facility; and (c) respondents might be a self-selected sample—for example, surveyors might be reaching only avid cyclists or other unusual groups.

To be able to generalize about the wider population—a critical criterion for this exercise—it was necessary to use a probability sample. Doing so, however, results in a more complicated sampling frame, usually resulting in lower rates of users who may walk or bike.

Sample size

The smaller the change to detect, the larger the sample size that is necessary to prove it significant. In the case of the pilot survey, sample size was largely influenced by costs. Relatively speaking, the sample was not exceptionally small. It needs to be recognized, however, that its size *may be* a factor contributing to the few statistically significant findings. However, sample size is not the only factor that determines the ability to detect effects with statistical significance (statistical power) (Forsyth, Krizek et al. 2011). Equally relevant are how much variation there is in the factors to be analyzed and how precise the tools are to assess these factors.

Random variation

It is in the nature of nonmotorized travel behavior that many relevant factors vary substantially between individuals, as well as by individuals over time. For example, the question "How long do you ride your bike on a typical day?" will result in a wide range of answers across people (because cycling habits are different) and also for the same individual, if one would ask the same individual several times over the course of a year (because cycling habits change with seasons, memory is influenced by recent experiences, etc.).

Measurement error
Measurement error is inherent to any instrument used to assess a factor. Often there is a trade-off between measurement error and other characteristics of an instrument. For example, to measure trip distance, self-reported questionnaires are inferior to GPS devices, but in contrast they can be sent out to large, representative population samples. As long as measurement error occurs randomly, it makes it harder to detect statistically significant findings (because it increases the variation in the observed data), but it will not result in spurious results (i.e., a change where in fact there was none). However, if error occurs systematically (bias), spurious results can be the result. A potential source of bias that is of main concern in the NTPP analysis is the comparability of the two study samples for the phase 1 and phase 2 surveys.

Sample characteristics
Related to the above point, the sampling strategies for the data collection were two-fold: a blanket coverage in each community and oversampling within key areas for NMT. Although conceptually designed to be equivalent, the possibility cannot be excluded that differences in the constitution of the two samples are responsible for underlying differences between the baseline and bookend survey, which may have resulted in unexpected changes or may have masked some changes where they would have been expected. Differences in average age (phase 1: 52y, phase 2: 58y) and proportion of women (phase 1: 53%, phase 2: 42%) are indications to that effect; however, sensitivity analyses showed that standardizing for the age and gender distribution did not affect the overall findings.

Several other factors of probably lesser relevance that may have contributed to the lack of consistent findings are briefly mentioned in the following text.

Geographical scale of intervention impact
Many of the intervention projects occurred in select corridors or for a specific population (e.g., a neighborhood). The larger the geographic unit (which in this case was considered to be the entire communities, for the most part), the less likely various populations are going to be affected. The effects of the local projects become diluted across the population samples, which makes them difficult to detect. Evaluation strategies and accompanying samples that are more geographically specific may needed to detect such effects.

Control community

In 2006, the research team recruited additional funding to collect data from a community that would, based on current information and the best judgment of experts at the time, be as suitable as possible to serve as a control community. After extensive deliberation, Spokane was selected (see 2006 report for rationale and considerations). One of the characteristics of quasi-experimental research—which this is—is that a "true" experiment is difficult to achieve. There are bound to be considerations that taint the control; it is just a matter of to what degree. For example, in 2005 Spokane had no ambitious plans for nonmotorized improvements.

As it turns out, several NMT projects were funded by the American Recovery and Reinvestment Act (ARRA) and realized during the study. For example, between 2006 and 2010, the city lengthened its system of class I facilities (shared-use path) by 10 miles (48.4 versus 59.2) and class II facilities (bike lanes) by 23 miles (53.0 versus 75.9). The city additionally hired its first bicycle and pedestrian planner during this time.

Possibility of lack of actual impact

There could be many reasons why the investments made may not have made a detectable impact on the use of nonmotorized travel based on the survey protocol in this exercise. One significant challenge with the setup of this program is that the effects of the various interventions did not have enough time to "mature" and have an impact. Some of the main infrastructure projects have only been in place for a very short period of time. However, it is well known that such long-term projects require some time to induce changes in travel behavior and for usage to build up.

3 Conclusions on population survey findings

It is important to point out that the inability to detect significant patterns of change is not synonymous to no change occurring. Given the variety of factors that may have contributed to the null-findings of this exercise, a key lesson to be learned from this pilot analysis is the need to improve, enhance, and make data-collection approaches more robust for NMT projects. Second, it is important to note that relevant considerations decided in a political context—such as the selection of a diverse set of communities and the implementation of a wide range of different policies—are often not working in favor of scientific evaluations. Future efforts should therefore carefully evaluate priorities and find a balance between exploring and implementing new policies and scientific evaluation of their effectiveness, and allocate funds accordingly.

4 Lessons learned

- Impacts of NMT interventions are difficult to demonstrate scientifically unless adequate means, resources, and time are allocated. In particular, pre–post evaluations that employ probability-based samples are extremely challenging without the availability of routinely collected data, such as regularly conducted household travel surveys and traffic counts.
- Political and administrative considerations in the planning stage of a pilot program may not always underscore the importance of scientific evaluation. It is therefore suggested that a platform for robust evaluation be considered a critical part of the planning process for such projects from their conception.
- Scientific evaluation of the effectiveness of pilot programs requires adequate resources and evaluation, and it is important that such resources be considered in the planning process from the onset of such a program.
- Scientific evidence on the effectiveness of NMT interventions is growing rapidly. Before engaging in evaluation efforts, it is highly advisable to consider existing evidence to avoid unnecessary replication of efforts.
- The scientific community is advised to invest in the development of data-collection and evaluation methodologies that are practice-friendly, cost-effective, and robust.
- The federal government should seek an active role in advancing and coordinating routine data-collection efforts for NMT to track and evaluate performance of NMT nationwide, as is the case for other modes.

5 Appendix

5.1 Related relevant documents

NTPP evaluation report, phase 1

Survey methodology report, phase 2,

Final Data, phase 2 (Click the "Final data set" link under Phase 2-2010 / Final Data Set)

Final Data Analysis, phase 2

5.2 Weighting procedures for the data

Weighting procedures differ for phase 1 and phase 2 because of slight differences in what was learned in phase 1 in terms of more efficiently administering the survey. Phase 1 consisted mainly of a two-stage survey—the self mailer and CATI—while in phase 2, only a CATI was conducted.

Phase 1

To adjust for oversampling when sending out self-mailer surveys, records were weighted using the inverse probability of being sampled. Sampling probability was calculated as "number of surveys sent out per census tract over total population of a census tract."

To adjust for opportunity sampling by reference mode in CATI, an additional weight was calculated as the inverse of "mode percent in the CATI sample (approximately 25%) over actual mode percent (as determined by self-mailer question q1c)." This weight is the same for each record of a particular reference mode and is applied wherever CATI variables are used to derive population averages (e.g., average time spent walking). For example, roughly one in four CATI are conducted with a "pedestrian" (reference mode walk), while based on the actual mode share, only roughly 10% of all CATI information should be provided by pedestrians. When calculating the overall population average (including drivers, etc.), a pedestrian's information will only receive 1/(25%/10%)=0.4 times the weight.

Phase 2

In phase 2, self-mailer and CATI were combined into one survey (CATI). For all questions, a simple sampling weight analogous to phase 1 was applied, based on "number of surveys sent out per census tract over total population of a census tract."

However, for a set of questions a reference mode was assigned based on what modes the subject had used recently. In cases of multiple modes used, reference mode was assigned based on a hierarchy—prioritizing rare modes such as cycling and transit—thereby affecting the proportion for each mode. For these questions a further weight was applied, analogous to the CATI weight in phase 1.

In some cases, additional weights were applied:

For mode share calculations for commutes to work or school (self-mailer questions 7 and 8), responses were weighted to assign equal weight to each respondent, discounting for multiple answers.

5.3 Primary output variables

Mode share

Mode share of a particular mode is usually defined as the proportion of all trips that is being taken by this particular mode.

"How many places did you visit yesterday?" (self-mailer question q1c), by mode, is used as the main variable to derive mode share (interpreting "place visited" as "a trip taken"). To assess changes between surveys and their statistical significance, mode share was first calculated on an individual level as "Number of places visited by a specific mode divided by the total number of places visited," then averaged and compared across surveys.

Three records with more than 30 car trips or more than 30 walk trips were excluded.

Alternative measures for mode share (and similar measures)

Commute mode share during past week to work and school (self-mailer questions q7 and q8), respectively, were calculated as "Number of commuters for specific mode over total number of commuters." Weights are applied to respondents who report multiple commute modes (e.g., a commuter reporting car and bike as commute modes was weighted to count as 0.5 car commuter and 0.5 bike commuter). (Note: the survey did not capture how many days each mode was used to commute.)

Measures of last use of a specific mode over various time periods (self-mailer question q2) were calculated as "Number of uses of a specific mode during specified time period over total uses of all modes during specified time period." To facilitate easier interpretation, the categorical variable (categories: past 7 days, past month, etc.) were converted into number of days since last use, using midpoints of categories (3.5 days, 15 days, etc.). The open-ended category "not used in the past year" was presumed to reflect, on average, a most recent use 540 days (1.5 years) ago.

Trip frequency

"How many places did you visit yesterday?" (self-mailer question q1c), by mode, is used as the main variable to derive trip frequency by mode (interpreting "place visited" as "a trip taken").

Records with more than 30 car trips or more than 30 walk trips are excluded.

Trip frequencies by mode are calculated per community and total sample as "Number of places visited by a specific mode per day."

In addition, separate variables assessed on how many days per week subjects walked at least 10 minutes per day (CATI a3), or biked (CATI a6), respectively.

Duration of walking and cycling

Days per week walking (CATI a3) or bicycling (CATI a6) were multiplied with duration of walking (CATI a4) and biking (CATI a7) per day, respectively, to derive duration per week.

Records of more than 120 minutes of walking or biking per day were capped at 120, to reduce influence of extreme values on overall means.

Duration is averaged both over all subjects, including those who did not report any days with more than 10 minutes of walking or biking, as well as across only those subjects who did report the two activities. (Note: in particular for biking, this makes a big difference.)

Note: original analysis (UMN evaluation report for phase 1) excluded durations of more than 60 minutes, which leads to a large number of excluded records (approx. 220).

Trip distance

Trip distance was determined in a GIS, based on origin, route, and destination information provided by subjects.

Bicycle trips shorter than 50m and longer than 14400m (9 miles) were excluded from the analysis, assuming that they reflected geocoding errors. Walking trips longer than 2 miles were excluded for the same reason. *(Such cut-off points were derived from UMN Evaluation Report for phase 1.)*

Trip distance was averaged by mode and community, as well as by mode across all communities (but not by communities, across all modes).

5.4 Secondary output variables

Access to and egress from transit

Two variables are available to measure access to (b10) and egress from (b11) transit. Access and egress modes are reported as percentages from all access and egress trips.

Trip purpose

Trip purpose was not asked for explicitly. Trips to destinations (q1a, q1c*) are interpreted as trips for transportation.

5.5 Calculated variables

Miles traveled

Distances traveled were derived by multiplying the

- mode specific average trip distance of each community with the
- mode specific average trip number per person in each community and with the
- number of people living in each community

(Trip distance * trip frequency * population)

5.6 Suggestions for benefit calculations

Based on the collected variables, benefits calculations are proposed using either established conversion factors or, in absence of such, by applying newly developed approaches.

Initially, the intention was to use observed changes in nonmotorized transport to calculate secondary benefits. However, due to the lack of consistently observable changes in core variables between phase 1 and phase 2 surveys, no benefit calculations were conducted.

In the following, the theoretical consideration that would have informed such benefit calculations is provided.

Vehicle-miles traveled avoided due to walking and bicycling

The survey instrument does not provide any direct information on whether, or to what extent, observed walking and bicycling resulted in a reduction of driving. In fact, it is often thought by researchers that such a phenomenon is impossible to reliably measure. Therefore, certain assumptions need to be taken before calculating avoided driving:

- Recreational walking or bicycling, defined as walking or bicycling with the primary purpose of exercise, should not be attributed with significant reductions in driving. (It is worth mentioning, however, that some recreational walking or bicycling may substitute for other forms of exercise, which might have required driving (i.e., to a workout facility). In such latter cases, it is conceivable that recreational walking or bicycling might be linked to decreased driving, although these amounts are extremely modest.)
- Any walking or bicycling to a destination could be considered as for utilitarian transportation purposes and therefore be attributed a reduction in driving (even if the destination is a recreational facility or a home-to-home loop, as above).
- The relationship between the distance of walking or biking trips and the distance of driving trips they replace is unclear and depends on various circumstances. There are at least three plausible scenarios, and their applicability depends strongly on their context (examples for when these scenarios are plausible are given in parentheses). The conducted surveys did not collect sufficient information to determine which substitution scenario best applies. Given the range in substitution factors, a 1:1 replacement may provide a reasonable midpoint:
    1. NMT trips are shorter than driving trips. For example, one mile of NMT replaces more than one and up to many miles of driving: Because the range of NMT is smaller than MT, subjects replace a driving trip to a relatively distant destination with a NMT trip to an equivalent destination closer by (example: shopping trip to a mall replaced by shopping trip to a local neighborhood store).
       Also thinkable, in particular for short trips: NMT provides a more direct, hence shorter route (example: downtown trip).
    2. NMT trips are of the same distance as driving trips (1:1 substitution): Travel mode does not affect choice of destination (example: commute).
    3. NMT trips are longer than driving trips: In particular, cyclists may be willing to go out of their (direct) way to encounter more favorable riding conditions (example: NMT on designated facilities, like trails).
- Similarly, the relationship between overall amount of walking and bicycling a respondent reports, and the overall amount of driving this reduces, is unclear. For example, a subject who engages in NMT may not have access to a car. There is no immediate effect on the level of driving. How to assess this case strongly depends on the context and purpose of the analysis:
    - If the question of interest is how much driving will be reduced if NMT is increased by a certain amount, NMT subjects who did not drive before should not be accounted for. (This approach was applied in the interim

report [based on whether subjects reported some driving (q5a), or access to vehicle (q6)?].)

- If the question of interest is how much it is worth that some people walk or bike, assuming that if they would not they would drive (even though they currently do not have the (immediate) possibility to do so), then their (all) NMT should be accounted for. (From an intermodal transportation system [or macro-economical] perspective, this approach is probably more adequate.)

Percentage reduction in auto travel

Conceptually, the observed change in auto mode share reflects this, although it is based on trips and therefore may over- or underestimate changes in volume (see assumptions for VMT reduction).

To receive the change based on VMT, the difference in the observed VMT (trip distance of reference trips by car * trips taken by car yesterday * 365 * driving age population) could be calculated as a percentage change between phase 1 and phase 2.

Alternatively, the calculated VMT avoided could be related to total VMT reported by other sources.

Gallons of fuel saved

Fuel savings can be calculated as VMT (vehicle miles of travel) avoided multiplied by average measure of fuel economy. For example, average fuel efficiency of the US passenger vehicle fleet is 20.2 miles per gallon (0.0495 gal/mi) as reported by the Bureau of Transportation Statistics (BTS); other conversion factors are available (Gotschi and Mills 2008; Gotschi 2011).

Avoided quantities of criteria pollutants and greenhouse gas emissions

Emission factors for CO_2 and criteria air pollutants can be applied to VMT avoided to calculate reductions in these emissions. For example, a commonly employed conversion factor assumes that 9.3kg CO2e are emitted for each gallon of gasoline (ICLEI, 2003); other conversion factors are available (see: http://www.epa.gov/oms/climate/420f05004.htm).

Increased physical activity

There is now ample evidence for the health benefits from physical activity in general (Physical Activity Guidelines Advisory Committee 2008), and from walking and bicycling in particular (Andersen, Schnohr et al. 2000; Matthews, Jurj et al. 2007; Hamer and Chida 2008). Based on these findings, minutes of physical activity from NMT can be derived based on assumptions of average assumptions for speed and subsequent health benefits; although, by many accounts, the body of knowledge and accuracy for such conversion factors is still in its infancy.

Safety

The conducted surveys do not provide the necessary information for safety assessments nor for improvements in safety over time. Safety evaluations require robust assessments of NMT (exposure variable) and reliable crash reports. Ideally, such data are available with high spatial resolutions, such as individual links of the road network. Collection efforts for both types of data require significant improvements before reliable safety assessments can be conducted. In places where such data are available, research shows that increases in NMT clearly reduce individuals' risk of injury, a phenomenon referred to as "safety in numbers" (Jacobsen 2003).

5.7 Final results

The final data and the final analysis of the analyses are compiled in a set of tables available as a separate document.

Final Data, phase 2 (Click the "Final data set" link under Phase 2-2010 / Final Data Set)

Final Data Analysis, phase 2

5.8 Development of community populations between phase 1 and phase 2

Community Populations

1=Baseline, 2=Bookend, d=Change (Bookend -Baseline)

Community	population_1	population_2	population_d	Percent change
Columbia	106582	119582.3	13000	12
Marin	223132	222857.7	-274	0
Minneapolis	348446	388238.7	39793	11
Sheboygan	112646	114222	1576	1
Spokane	400049	455204	55155	14
all regions	1.00E+06	1107750	73675	7

6 References

Andersen, L. B., P. Schnohr, et al. (2000). "All-cause Mortality Associated with Physical Activity During Leisure Time, Work, Sports, and Cycling to Work." *Arch Intern Med* **160**(11): 1621-1628.

Cavill, N., S. Kahlmeier, et al. (2009). "Economic Analyses of Transport Infrastructure and Policies Including Health Effects Related to Cycling and Walking: A Systematic Review." *Transport Policy* **15**(5): 291-304.

City of Portland (2008). "Portland Bicycle Counts 2008." Accessed October 2008, http://www.portlandonline.com/transportation/index.cfm?c=44597&a=217489.

City of Portland Office of the Auditor (2007). *Service Efforts & Accomplishments Report 2006-7*.

City of Portland Office of Transportation (2010). "Portland Bicycle Plan for 2030." Accessed July 27, 2010, www.portlandonline.com/transportation/index.cfm?c=44597.

Forsyth, A., K. J. Krizek, et al. (2011). "Sampling Issues in Nonmotorized Travel Surveys: The PABS (Pedestrian and Bicycling 1 Survey) Approach." Presented at Meetings of the Transportation Research Board.

Gotschi, T. (2011). "Costs and Benefits of Bicycling Investments in Portland, Oregon." *Journal of Physical Activity and Health* **8**(Suppl 1): S49-S58.

Gotschi, T., and K. Mills (2008). *Active Transportation for America — The Case for Increased Federal Investment in Bicycling and Walking*. Washington, D.C.: Rails-to-Trails Conservancy.

Hamer, M., and Y. Chida (2008). "Walking and Primary Prevention: A Meta-analysis of Prospective Cohort Studies." *Br J Sports Med* **42**(4): 238-243.

ICLEI (2003). Clean Air and Climate Protection Software (CACPS); ICLEI Local Communities for Sustainability, 2003. Accessed May 2011 http://www.4cleanair.org/InnovationDetails.asp?innoid=1.

Jacobsen, P. L. (2003). "Safety in Numbers: More Walkers and Bicyclists, Safer Walking and Bicycling." *Injury Prevention* **9**(3): 205-209.

Krizek, K., S. Handy, et al. (2009). "Explaining Changes in Walking and Bicycling Behavior: Challenges for Transportation Research." *Environment and Planning B: Planning and Design* **36**(4): 725-740.

Levinson, D., R. Karamalaputi, et al. (2003). *If They Come, Will You Build It? Urban Transportation Growth Models.* St.Paul, MN: Department of Civil Engineering, University of Minnesota.

Matthews, C. E., A. L. Jurj, et al. (2007). "Influence of Exercise, Walking, Cycling, and Overall Nonexercise Physical Activity on Mortality in Chinese Women." *Am. J. Epidemiol.* **165**(12): 1343-1350.

Physical Activity Guidelines Advisory Committee (2008). *Physical Activity Guidelines Advisory Committee Report 2008*. Washington, D.C.: U.S. Department of Health and Human Services.

Wang, G., C. A. Macera, et al. (2005). "A Cost-Benefit Analysis of Physical Activity Using Bike/Pedestrian Trails." *Health Promotion Practice* **6** (2): 174-179.

WHO. (2008). "Health Economic Assessment Tool for cycling." from http://www.euro.who.int/HEAT.