EUROPEAN COMMISSION
Seventh Framework Programme

Capitalising CIVITAS Knowledge and experience

CIVITAS CAPITAL

# WP1 Thematic Groups

# How-to Guide on Long-term Evaluation

D 1.4

Authors:
Tom Rye, Anders Wretstrand

Lund University
November 2014

**Table of Contents**

# 1 Introduction to the project and this document

## 1.1 CAPITAL's place in the CIVITAS Family

CIVITAS CAPITAL is one of the two support actions that are currently running under the CIVITAS Initiative. It builds on previous actions such as VANGUARD and CATALYST and offers its support to DYN@MO and 2MOVE2. It helps CIVITAS to develop a strong identity towards Horizon 2020. Next to that, CAPITAL seeks synergies with CIVITAS WIKI, with which we share our communication channels, thematic cooperation activities and online collaboration platform.

## 1.2 Mission statement

The mission of CIVITAS CAPITAL is to …

> *"…contribute significantly to the goals of the EU's Transport White Paper by capitalising systematically on the results of CIVITAS and creating an effective "value chain" for urban mobility innovation. CAPITAL will initiate and support a mainstreaming process of CIVITAS principles based on a strengthened community of stakeholders. CAPITAL will help CIVITAS to build the bridge towards a more advanced identity within Horizon 2020. It will help to create a more structured link with large-scale deployment in support of Transport White Paper goals."*

## 1.3 Goals and objectives of Work Package 1: Thematic Groups and the purpose of this document

This work package will:

- contribute to the project's Best Practice Clearing House by capitalising on the work of the Thematic Groups for the eight CIVITAS measure categories and for other topics;
- contribute to the project's Knowledge Centre, stimulating programme-level knowledge transfer, dissemination and long-term evaluation for inputs to CIVITAS Advisory Groups and therefore further take-up of measures;
- provide guidance from the point of view of practitioners view to developing policy packages and replicating successful CIVITAS experiences.

As part of its work to achieve the second bullet point, CAPITAL will build on the experiences of the CATALIST project in long-term evaluation of measures. CATALIST carried out the very first long-term evaluation exercise within CIVITAS between 2007 and 2012. In CAPITAL, the objective for long-term evaluation is to find a maximum of two measures per TG that can be subject to long-term evaluation. To support this, CAPITAL will encourage cities to apply for additional funding for long-term evaluation exercises from the Activity Fund in Thematic Group areas which are relevant to the take-up of new activities and / or activities in the national networks.

This “How to” Guide on long-term evaluation is a key part of this process as it provides the basis for training and assisting TG members, and others, especially former measure leaders, in applying long term evaluation to their chosen measures and delivering the evaluation results.

This document first provides a definition of long-term evaluation, and then some advice on how to do it. It then presents some results of previous long term evaluations of transport measures, mainly from outside CIVITAS projects, in order to show the benefit of carrying out long-term evaluation.

# 2 Evaluation – on different terms

Is there a difference between short-term and long-term evaluation, and if so what is the difference? Well, the short answer to that question is that while short term stretches over a short time period, the long term perspective could be 5-10 years or even more. In the CIVITAS Initiative context, it could be said that (short-term) evaluation activities are carried out during the project life span, often 3-4 years, while long-term evaluation activities are taken up later after the project time in order to do follow up studies.

Short-term evaluation generally covers one year or year or duration of project funding and uses before and after data. Long-term evaluation can include projecting impacts into future through forecasting and scenario-building. It can involve time series data for schemes or measures over years, using running, ongoing surveys or planned, repeated long-term effect surveys.

Long-term evaluation could thus seek to answer whether the long-term impacts of measures have been different from short-term impacts. Or, if the CIVITAS pilot measures have been up-scaled as an effect of the relative successful pilot implementation, and eventually have come to cover a larger geographical area or system, then how much greater proportionally has the impact been?.

Do short and long term evaluation lead to different results? It is reasonable to assume that some structural effects will show. In some cases there may be no impact evidence in the short term, but detectable positive impacts on long term, consistent with much of the literature on price elasticity of demand, for example, which shows that effects of price changes for public transport or fuel are greater in the long term than in the short term. In other cases there will be impacts in the short term, but because of poor measure maintenance, “rebound” effects (where people get used to a measure such as, for example, road user charging), or changes in background conditions, the impacts decline. In both cases the processes (the stories) are important to capture; what has actually happened over time since the measure or scheme was launched?

Another difference may be related to the actual scale of the project and its objectives. Long term evaluation most likely will come into play in case of large-scale, long-term, multi-site comparative designs compared to short evaluations of a single measure in a city. For either ends of the time or size scales, there are basically two fields of assessment – the impact and process evaluation. The characteristics of these fields will be briefly outlined below. For further references, see the CIVITAS II Publication *Evaluation Matters* (Technical University of Berlin, 2013).

## 2.1 Impact and process evaluation

Impact evaluation seeks to describe the effects of the measure’s implementation in comparison with the situation before the implementation. The impacts that usually are focused on in the evaluation are strongly related to the objectives of the measure. The impact is not the actual new scheme itself (the output) but the outcome: e.g. the impact it has on people’s mobility or the urban environment.

To be able to assess the outcome, the objectives have to be expressed as measurable indicators. Doing the long-term evaluation follow-up requires access to or at least knowledge of previous evaluation elements and activities, indicator data collection methods and analyses, so that short and long term impacts can be compared. Impact evaluation is often based on quantitative data; time series may be available and annually updated even after the project finished years ago. An example of the latter is fuel consumption data for alternatively fuelled buses, as most operators will collect this data regularly and retain it over time.

Process evaluation focuses on the means and procedures by which a measure is implemented; it tells the story of planning, implementing and operating the new scheme, technology or infrastructure. Hence, it begins during project development and continues throughout the life of the project. Its intent is to assess all project activities, negative and positive factors which are influencing the measure implementation process and thus provide information to monitor and improve the project, as well as information and guidance to followers who may wish to emulate the project.

Doing the long-term evaluation follow-up, access or at least knowledge of previous process evaluation elements and activities will allow for simplified procedures and asking the right persons; qualitative interviews are very common methods in process evaluations.

## 2.2 General evaluation issues

Evaluation is not always simple and clear cut. Each measure has one or more objectives, and there may be a combination, a bundle, of measures. Objectives may also be very overarching: a large modal shift between private car and bus, a livelier and cleaner city centre, coordinated urban freight. Then, these have to be operationalized into indicators that are quantifiable, measurable, and still clearly linked to the objectives. Each objective could relate to several measures, whereas each measure should be linked to several indicators.

Figures 2.1a-d below show two things: 1) it is important to be able to control for (or at least estimate the impact of) other confounding factors; 2) the impact of the measure itself as well as other factors may vary over time. One important aspect of long-term evaluation is thus to be able to assess changes in the context. What has changed since the measure was implemented and how can we estimate these changes and the impacts they may have on what we intend to measure?

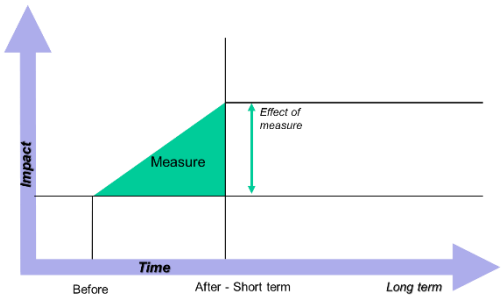


Figure 2.1a: Short term evaluation shows significant effect of measure

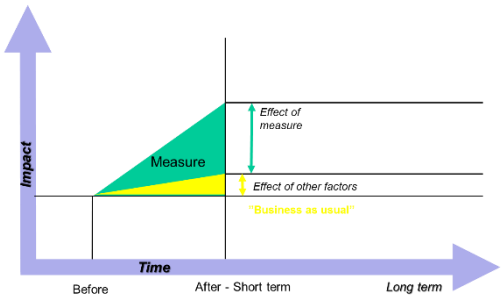


Figure 2.1b: While considering the do-nothing scenario effects, the effects are significant but not solely caused by the measure

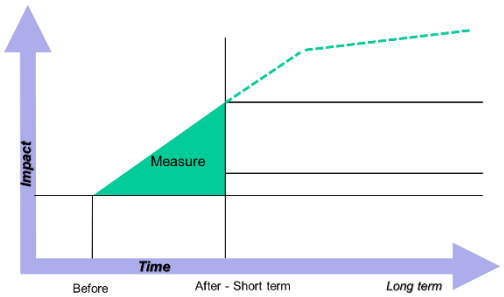


Figure 2.1c: on longer term, the impact increases, but the increase rate diminishes over time.

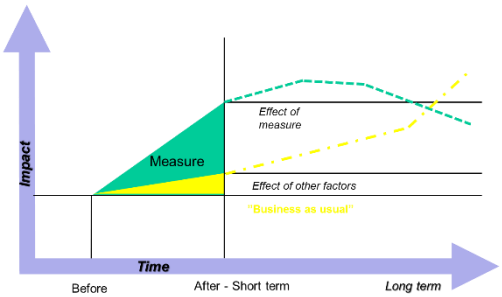


Figure 2.1d: Here the impact of the measure is in fact starting to decline, whereas impacts of other factor increase

## 2.3 Availability of documents

We assume that the measures or schemes of interest have all gone through the process of planning, implementation and (short term) evaluation. Then, it is recommended to look for the following steps that have been taken towards project realisation:

- Clearly defined objectives?
- Which target groups?
- Other measures related to the same objectives?
- Which factors were chosen as indicators?
- Where targets for success set? For certain areas or population of users?
- Were these targets met in the short term?
- Type of data collection and study design?
- Clear results, incl. various analyses (Cost Benefit Analysis (CBA), up-scaling, transferability etc.)?

If it concerns a previous CIVITAS Initiative project, or is co- funded through other European or National funding sources, it is likely that this information will be publicly available, but parts may be difficult to assess. It may for instance be possible to obtain information of a specific measure, but as indicated in figures 2.1a-d, the “true” short term effect maybe more difficult to distinguish, as several measures may interact. Therefore, it is recommended to get as much information as possible of all the measures that were included in the program or project.

## 2.4 Evaluation design

While collecting information about the measure and previous evaluation, it is important to clarify the evaluation design. Referring to the impact-time charts above in Figures 2.1a-d, the study design tells a lot how certain one can be of the “true” measure impacts.

In general, the evaluation design is a plan for collecting and analysing evidence that the measure will have the impact it purports to have. The earlier choice for a particular design is frequently influenced by the need to compromise. The more certain the answers, the more costly the evaluation, and vice versa.

Below in Figures 2.2a-f are some examples that might characterize the situation while entering on long-term follow-up studies. If a case/control design – the theoretically preferred research design, which allows the effect of the measure to be isolated from the effect of background factors such as the wider economy, fuel price increases and so on - has been used, it is likely to be of a quite limited scale. Rather, we might be looking at measures with baseline partly or totally lacking. It is also possible that situation 2.2f occurs; the objective of the long-term evaluation refers to a certain measure, but for some reason the main indicator connected to the objective has not been measured properly and thus, the results are of no use. If this happens to be the case, it is not recommended to do a long-term impact evaluation at all (process evaluation would still be possible to do).

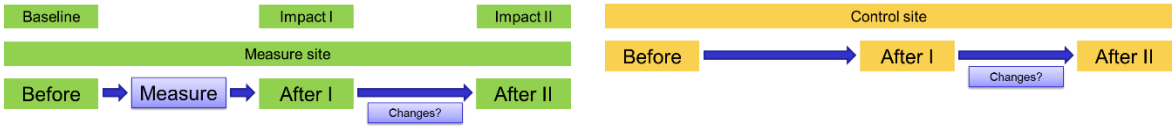


Figure 2.2a: Evaluation with control group/site, baseline data for case/control, “After I” are short term impacts and “After II” long term impacts


Figure 2.2b: Evaluation with control group/site, baseline data for case only

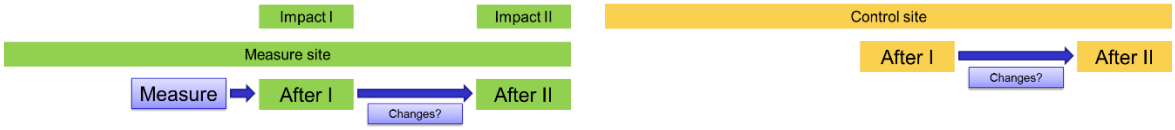


Figure 2.2c: Evaluation with control group/site, no baseline data available


Figure 2.2d: Evaluation of case only, with baseline data


Figure 2.2e: Evaluation of case only, no baseline data

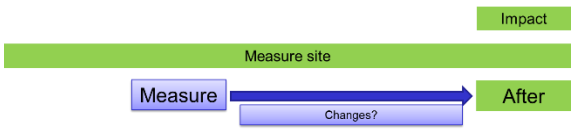


Figure 2.2f: Evaluation of case only, neither baseline nor short term impact data

## 2.5 Scenarios and forecasting

Forecasting can be used to provide a prediction or estimate of the impacts of a measure, or of the city's transport system without the measure. It is therefore useful as a means of establishing the business-as-usual scenario, but also the scenario with the planned measures, or with a different set of measures, as a comparator. Once actual before and after data are available, these can be compared with the predictions to see how accurate these were. Scenarios can also be used in measure selection to help to choose between packages of measures – different scenarios paint different pictures of the future with a given package of measures.

## 2.6 Data collection and survey techniques

Data collection encompasses a wide variety of methods, data sources and units of data elements. Looking at, for example, the CIVITAS Core Indicators, corresponding data are either derived or measured. Data could be physical units (e.g. vehicles or pedestrians counted as they pass a given point on a street), economic data, or people's revealed preferences and behaviour collected through survey instruments.

Sometimes the key issue in the longer term is to repeat the short-term evaluation, and in this case it is key to follow the earlier procedures. Data sources may be available as continuous data series such as operational data or biannual surveys. Then, it is strongly advised to use these ongoing data sets, as it also allows for later follow-ups and monitoring through continuous time series.

However, the long-term evaluation perhaps aims to look at impacts and processes with a "fresh eye" and go beyond earlier designs. Therefore, some guidelines for conduction a survey are summarized below. For further reading, refer to *Evaluation Matters* (Technical University of Berlin, 2013).

The survey process contains of the following steps:

- Define survey purpose and be very clear as to the key facts that the survey is intended to obtain.
- Undertake preliminary planning:
    - Collect background information.
    - Design sampling.
- Select survey method.
- Design survey instrument.
- Conduct pilot.
- Implement survey.

At the preliminary planning stage the user faces the choice of doing a quantitative or qualitative study. The purpose of impact evaluation generally is to obtain information from a broad cross section of users - a population that one would like to describe through a sample.

A qualitative approach would be much more appropriate if the key aim is to gather “softer” more explanatory data about *why* measures have had the observed effect.

However, to be able to say something meaningful about the long-term impacts of a measure solely based on personal interviews or focus groups is likely to be challenging. Therefore, from now on only quantitative designs are referred to as the main components of impact evaluation. But for the analyses of processes in both the short and long-term, a more qualitative approach is recommended.

The use of available guidelines for survey design and sample selection, such as the excellent publication *Evaluation Matters*, based on experience in previous CIVITAS projects (Technical University of Berlin, 2013), is also recommended. If possible, use similar designs as previously used in the short-term evaluation, but do not assume that chosen designs are by definition the best possible solutions – review them against best practice.

If the comparison between short-term and long-term impact is crucial, changes to the survey design compared to the method used in the short term should be avoided as far as possible. If the long term impact, or in fact the circumstances when the survey is conducted is more important, the survey validity and reliability should be key. For example, if the previous modal split survey had obvious flaws, there is no reason to repeat a poor design unless comparability between surveys is most important.

# 3 Long-term evaluation – step-by-step

This section summarises the step-by-step approach of long-term impact and process evaluation. This will require that research is undertaken: to be able to take these steps, background data such as programme or measure descriptions as well as evaluation reports need to be collected and analysed. The following steps are also summarised in Figure 3.1.

## 3.1 Impact evaluation

### 3.1.1 Objectives

Review policy/measure objectives! Impact evaluation illustrates changes which are attributed to a policy or measure which aimed and were implemented to reach specific objectives. To start with, we therefore have to review the policy or measure objectives? Which were they?

If possible, in order to clarify break down the policies/objectives into discrete parts. Maybe there are overarching as well as specific objectives, where the latter may be more relevant when it comes to short-term evaluation efforts.

### 3.1.2 Cause / effect

Reflect on cause and effect relations! The outcome of a measure can always be caused by a variety of effects which need to be considered, because some impacts are often indirect, with several steps between an activity and its eventual impact. Try to map possible factors that could contribute to or negatively affect the desired state (objectives). Which of these factors seem most suitable to measure?

How did the previous evaluation consider these aspects? Moreover, can we expect specific long-term effects, such as behavioural adaptations and changes (people moving, buying/selling private cars)?

### 3.1.3 Indicators used

Check short-term evaluation indicators! How are they described, and in which units? They should clearly relate to and measure the discrete objectives. Which type of data collection and survey instruments were used? How easy is it to obtain the raw data from earlier rounds of data collection?

### 3.1.4 Methods used

Analyse or review short-term impact evaluation! Do the results reveal solid rigorous methods, or can poor results be a consequence of poor design? Good methods could be continued and repeated, but poor methods should be considered to be left out and replaced.

### 3.1.5 Indicator revisions

Revise and refine indicator descriptions! One reason for poor evaluation is the vague definition of an indicator, or that it is invalid. Good results often come with simple and clear-cut indicators that can easily be followed through reliable data sets. It is recommended to keep using the same indicators for long term follow up. However, it is convenient to make

use of ongoing surveys, and these may lead to a slight change in the indicator. Therefore, some indicators may have to be revised in order to be measured at all.

### 3.1.6 Collect new data

Collect new data for long-term evaluation, including changes in the background situation! In line with the previous section, align with short-term evaluation to continue the time series. Use the same/similar designs, sources, formats, sample strategies. Acknowledge the background and contextual changes and the other confounding factors may have had less or more impact on cause and effect.

### 3.1.7 Analyse data

Use proper methods for analysis, based on guidelines! Read the existing evaluation reports together with general evaluation guidelines. What lessons could be learned from short-term evaluation? What seems to be specifically context dependent? Are these underlying factors or structures either acting as barriers or drivers for the outcomes and impacts?

### 3.1.8 Interpret data

Are some indicators influenced by other measures? Do they interact as bundles? Start with the short-term analysis. How did the evaluators interpret their results? Maybe, the short-term evaluation was conducted far too close to the start of implementation, and it was therefore not yet possible to detect an impact.

Related to the cause and effect chain, it is imperative to acknowledge larger changes in the infrastructure or legal/fiscal framework. E.g., a long-term change in modal split due to real time information availability (a short-term evaluation indicator) is more likely caused by inner city access restrictions or large network changes. It is important to be able to address the relative importance of measures, as known from other studies.

## 3.2 Process evaluation

### 3.2.1 Objective data

Evaluate existing documents! Use the short-term process evaluation as a template for document analysis. Describe the changes in the legal framework in terms of planning, infrastructure and governance. If new policies have come into place, can they be detected in the local context and have they resulted in concrete changes on the ground?

It is of particular interest to follow up and see whether lessons learnt in the short term and hopefully documented in the first process evaluation led to revised strategies or upscaling of measures in the longer term. If earlier results were less positive, and measures have been down-scaled, it is probably unlikely that a long-term process evaluation will have been of such importance in the longer term implementation of the measure.

### 3.2.2 Subjective data

Ask stakeholders to do a self-assessment! Key stakeholders who have been working for a long time in the city administration may be very good sources of information (and provide several of the data needed for evaluation). Their expertise and opinion is a prerequisite for

interesting qualitative results. Key participants will be able to tell the story after implementation, about how the internal discussions and decision-making processes will affect future urban planning.

Qualitative interview data could either be collected through individuals, groups or preferably both – slightly different information comes from groups and individuals. Groups reveal the common discourses, whereas individuals focus more on details and facts.

| Steps | Remarks |
|---|---|
| **Impact evaluation** | |
| Review policy/measure objectives | Break down the policies/objectives into discrete parts |
| Reflect on cause and effect relations | Possible long-term behavioural adaptations and changes? |
| Check short-term evaluation indicators | They should match and measure the discrete objectives |
| Analyse or review short-term impact evaluation | Good methods to be continued, poor methods to be left out |
| Revise and refine indicator descriptions | Make use of ongoing surveys, revise some indicators if needed |
| Collect new data for LT evaluation, including changes in background situation | Align with short term evaluation to continue the time series |
| Analyse data | Use proper methods for analysis, based on guidelines |
| Interpret data | Are some indicators influenced by other measures? Bundles? |
| **Process evaluation** | |
| Evaluation of existing documents | New policies in place, upscaling of measures, new funding, etc. |
| Self-assessment of stakeholders | Interviews, workshops: reported failure, success, lessons learned |

*Figure 3.1 – Steps in long term evaluation*

## 3.3 Types of measures that should be selected for long-term evaluation in CIVITAS CAPITAL – how to choose them.

Within CIVITAS CAPITAL the intention is that certain measures implemented and evaluated in previous CIVITAS city projects can be identified and subject to a further round of evaluation. The purpose of this will be to measure any long-term impacts of the measures compared to those observed in the short term, and to understand the reasons for the changes in impacts observed. Suitable measures to select for this round of long-term evaluation are not limited to any particular area of sustainable transport but are rather those that:

- Have robust impact evaluation data from the evaluation during the project period.
- Had some impact measured.
- Have good process evaluation data.
- Do not require very complicated data gathering.
- Have a measure leader or other staff member who has some knowledge of the measure and interest in seeing it subject to long term evaluation.
- Have well defined clear objectives and indicators.

# 4 Some long-term evaluation results

The purpose of this section of the report is to present some actual examples of the results of long-term evaluation. These show the need to collect over some period of time data that can be seen to relate to the objectives of the transport measure of interest. However, as is the case when we consider the review of the economic impacts of transport investment, this data may be collected by another agency external to the project itself. Presented here are long-term evaluations of:

- City transport policies and their impacts on travel patterns in each city.
- Cycling policies in four large German cities.
- Mobility management measures in Australia.
- A small number of large transport investments – principally evaluations of their out-turn costs and benefit to cost ratios, but also evaluation in one case of their impacts.
- The link between transport investment (high-speed rail and motorway construction) and economic growth in EU member states.

This is a disparate list mainly because it is difficult to obtain long-term evaluation data. However, it is also an interesting set of cases where each case in its own way shows the value of gathering evaluation data over the long term.

## 4.1 City transport policies

When cities implement packages of sustainable transport measures, it can take some time for the results to take effect, as people will not immediately change their travel patterns in response to changes in new transport measures. A small number of cities monitor the travel patterns of their residents or of travellers in their area in order to see whether their transport policies and measures are having an effect. Some results of such monitoring are presented here, together with a short summary of the transport measures implemented, and a link to more information.

In terms of the “ideal” of long-term evaluation, most of the cities whose results are presented do not in the same document present data about, for example, economic activity in their area, or try to compare their results with those from other cities, which would make it easier to establish whether the changes observed are a result of the transport measures, or of some other factor. An exception to this is data presented below from the City of Nottingham. A further issue is that, even when the data are available on what has been implemented and the resulting changes in travel patterns, information is often lacking on *how* the measures were implemented (process evaluation). Nonetheless, these results are valuable as such information on long-term changes related to a city’s transport policies is rarely available in one place. Observing such results would not have been possible, of course, without a long term evaluation process in place, which underlines the value of carrying out evaluation over a period longer than just the project lifetime.

### 4.1.1 Burgos, Spain

Burgos is an historic city of just under 180,000 people in the north of Spain. It adopted its SUMP in 2005 during its participation, and as a specific measure for which it received EU funding, in the first round of the EU programme CIVITAS. The SUMP contained objectives and targets related to reductions in congestion and pollution and improvements in quality of life, especially in the city centre. Measures to achieve these objectives included:

- Pedestrianisation of the city centre, with access controls controlled by ICT.
- Construction of 50 km of bicycle tracks (from zero).
- Re-design of the bus network and introduction of new vehicles.
- Campaigns and awareness-raising with citizens.
- Parking management.

Burgos' monitoring data shows that from 2003 to 2009 it cut the proportion of trips in the city made by car from 35% to 27%, whilst increasing the use of public transport and cycling – but in part at the cost of walking, as some pedestrians switched to bike and public transport (for details see http://www.civitas.eu/index.php?id=66&sel_menu=35&city_id=3, accessed 12 May 2013).

### 4.1.2 York, Nottingham and Edinburgh, UK

The information in this section is sourced from these three UK cities' LTP (SUMP) websites and monitoring reports available on those sites (City of Nottingham, 2008; City of York, 2008; City of Edinburgh, 2007). They were chosen because these reports are of high quality and demonstrate some long-term success apparently resulting from the measures that they implemented.

All three cities are historic centres but Nottingham and Edinburgh are much larger (population of approximately 500,000 people compared to around 100,000 in York) and, while Nottingham has historically been quite industrial, Edinburgh and York are less so. In their SUMPs, these three cities used broadly similar packages of measures to try to reduce car use in order to tackle congestion and pollution; these packages consist of:

- Restraint-based parking policy;
- Park and ride;
- High quality buses on simplified network with simple fares structure (and a single tram line in Nottingham);
- Pedestrianised, high quality city centre (Nottingham and York);
- Reduction in road capacity in city centre;
- Widespread implementation of 30kph zones in residential areas;
- Improvements for cyclists; and
- Linking land use planning and sustainable mobility by, for example, ensuring that new developments are linked into public transport and the cycling network.

Results from Nottingham show traffic levels on its main roads stable from 2001-2008, compared to a growth of 12% on comparable roads in urban England (see Figure 3.1); and a total public transport trips up from 67m in 2003 to 75m in 2008. In York, bus passenger

numbers grew 50% from 2001 to 2008, and 25% in Edinburgh from 1998 to 2005. All three cities over-achieved against national targets in terms of reduction in people killed and seriously injured in traffic collisions.

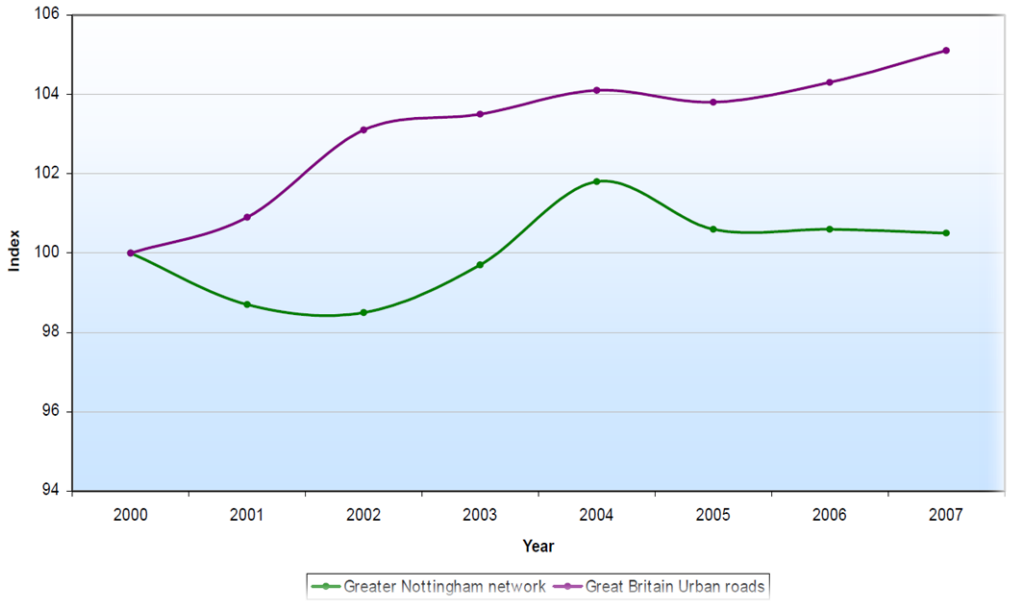


*Figure 4.1 – traffic grown in Greater Nottingham and comparable roads in Great Britain 2000-2007 (source: City of Nottingham, 2008)*

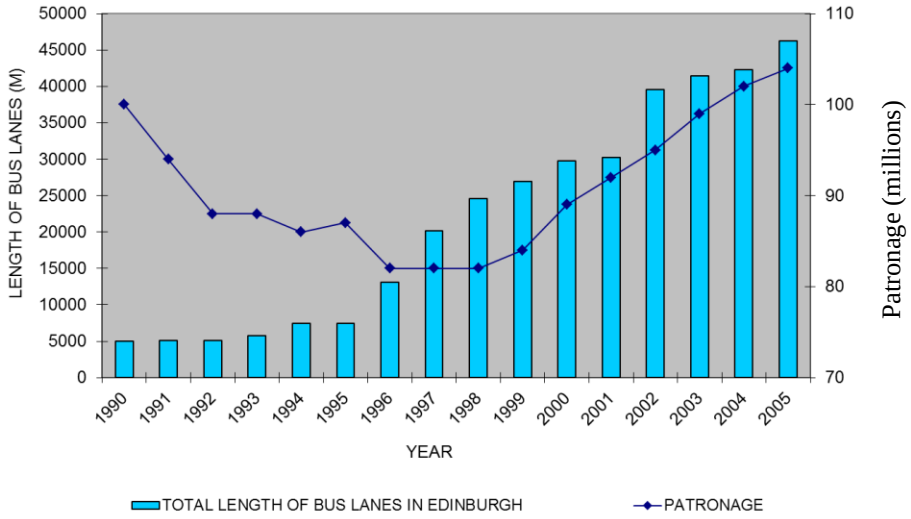


*Figure 4.2 – bus lanes and bus patronage in Edinburgh, 1990-2005 (Source: City of Edinburgh, 2007)*

### 4.1.3 Strasbourg, France

Moving to France, some mode share data are available from Strasbourg to show the impact on travel behaviour of the measures that it implemented as part of its PDU (SUMP). Again

through a package of measures not dissimilar to those for the UK cities described above, but with the very important addition in the French case of a very successful tram network linked to park and ride, the city has made significant progress towards a much more sustainable modal split, with an almost 14% reduction in car use between 1997 and 2009, from 53% to 46% of trips.

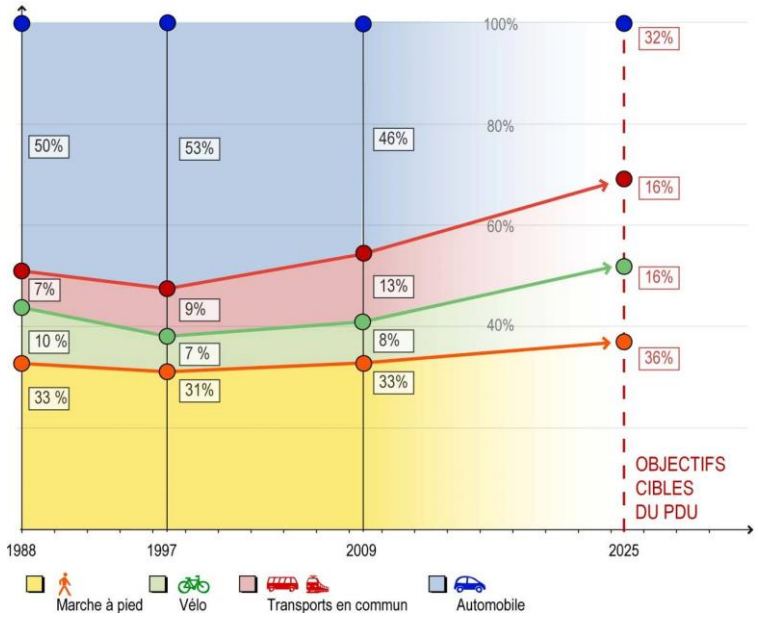


*Figure 4.3 – PDU outcomes and targets, Strasbourg (Source: Communite Urbain Strasbourg, 2012)*

This brief review of cases of a few cities' transport policies over time has shown that they can deliver changes in travel patterns at the city level. Without long-term evaluation they would have been unable to demonstrate this. As noted in the introduction, however, there is a lack of control data (such as the graph of traffic growth in other parts of England shown in Fig 3.1) which makes it difficult to confidently attribute the changes in travel patterns to the transport measures implemented. There is also often a lack of detailed monitoring of what is actually implemented, so it is problematic to understand the scale of the transport measures that cities have employed. Finally, there is a lack of process evaluation to understand how measures were implemented. Long-term data on outcomes in terms of travel patterns is therefore only part of the picture!

## 4.2 The long-term impacts of cycling policies in German cities

Lanzendorf and Busch-Geertsma (2014) provide evidence of what they call "a cycling boom" in large German cities. To do this they use German national travel survey data from 2002 and 2008 to show the share of all trips made by bike, plus any city level data from earlier years, also showing modal share. They compare the cities of Berlin, Frankfurt am Main, Munich and Hamburg, as well as data for all large German cities (those with 500,000 people or more) and for Germany as a whole. The first three of the four named cities were chosen

because they have had cycling strategies complemented by measures to encourage cycling in place since the early 2000s at least, whilst Hamburg has only recently adopted its cycling strategy. Between 2002 and 2008 the respective percentage increases in cycle mode share were 37.2% for Berlin, 35.9% for Frankfurt, 34.1% for Munich, 30.7% for all large German cities, 15.0% for Germany as a whole but only 2.7% for Hamburg – albeit on a high base for this latter city.

Using document analysis and expert interviews, the authors tell the story of bicycle promotion in each of the four cities and then try to piece together the *processes* that have occurred in the first three cities that they argue have helped to achieve these well above average increases in mode share for cycling. The key reasons that they point to are:

- The formal adoption by the city council (politicians) of a cycling strategy including ambitious targets.
- A minimum level of financing guaranteed to support the implementation of the cycling strategy.
- As a result of the above two factors, the improvement of cycling facilities.
- The combination of "hard" infrastructural measures with "soft" marketing and communication measures.

In this case of long-term evaluation, the work has been carried out by academics using nationally collected travel survey data (although note that cities in Germany often pay to increase the sample size of this survey so that results are reliable at the level of their city and can be used for transport planning purposes); plus a process evaluation in the four named cities. This kind of long-term evaluation provides important guidance to other cities that are seeking to find ways to increase their mode share for cycling.

## 4.3 Revisiting ex-ante evaluations of transport investments after schemes open

Another form of long-term evaluation is when the actual results of transport measures are compared with the predictions that were made before they were built, and on the basis of which the investment in the measure was justified. This relates primarily to large infrastructure investments such as new road and rail schemes where a cost-benefit analysis (CBA) is produced to compare the future benefits and costs of such schemes over a 30-60 year time horizon. Normally, the predicted benefits should exceed the costs for the scheme to be funded. Obviously some form of model must be used in order to predict the usage of the scheme 30-60 years into the future but these models' predictions are subject to inaccuracy. Nonetheless, relatively few researchers have gone back after the scheme has opened to compare model predictions with the actual project outturn.

Nicolaisen and Driscoll (2014) reviewed the studies that are available and their summary of the findings are presented in Figure 4.4, below. A negative figure in the column "Mean" indicates that usage of the scheme was over-predicted (actual use was much lower than forecast), and a positive figure the opposite, with the percentage showing by how much. The standard deviation shows the very wide variability in the forecasts for individual

schemes, compared to the mean across each study. It can be seen that there is a tendency for usage of rail schemes to be over-predicted and for road schemes to be under-predicted, with a very high standard deviation in almost all cases.

| Author(s) | Sample[a] | Mean | Standard deviation |
|---|---|---|---|
| Mackinder and Evans (1981) | Road: 44 | −7%[b] | N/A |
| NAO (1988) | Road: 128 | +8% | 43 |
| Pickrell (1990) | Rail: 9 | −65% | 17 |
| Flyvbjerg et al. (2006) | Road: 183 | +10% | 44 |
| | Rail: 27 | −40% | 52 |
| DoT (2007) | Rail: 19 | −37% | 31 |
| DoT (2008) | Rail: 18 | −16% | 59 |
| Bain (2009) | Toll: 104 | −23% | 26 |
| Button et al. (2010) | Rail: 44[c] | −21% | 58 |
| Parthasarathi and Levinson (2010) | Road: 108 | +6% | 41 |
| HA (2011) | Road: 62 | +3 | 21 |
| Welde and Odeck (2011) | Toll: 25 | −3% | 22 |
| | Road: 25 | +19% | 21 |
| Nicolaisen (2012) | Road: 146 | +11% | 35 |
| | Rail: 31 | −18% | 33 |

*Figure 4.4. Comparison of means and standard deviations for demand forecast inaccuracy – from Nicolaisen and Driscoll, 2014*

A specific example of a single scheme is presented by Odgers and Low (2010). The Citylink is a 22km urban motorway project that was opened in Melbourne (Australia) in 2000. Its primary objective was to reduce congestion and travel time. The authors went back to the project one and ten years after opening and gathered data on travel time and vehicle speeds, and found that on almost all counts this period saw an increase in network travel time and reduction in vehicle speeds. Figure 4.5 shows projected and actual impacts on citywide road travel time and speeds. The situation could potentially have been worse without the new motorway, but the long term evaluation shows that the model used was unable to make an accurate prediction and this led to an over-prediction of its benefits.

| Expected | | | Actual |
|---|---|---|---|
| Variable | Year | Specific value projected | Actual value reported [6] |
| 1. Average travel speed inner Melbourne; all roads whole of day (km /hr)[1] | 2001 | 36.1 | 33.8 |
| 2. Average travel speed across urban road network (km/ hr) [2] | 2001 | 44.3 | 42.6 (2000-01) |
| 3. Total daily vehicle hours of travel on the urban road network (000 hrs) [3] | 2001 | 1929.8 | 1991.8 (1999-00) 2087.2 (2000-01) |
| 4. Total daily vehicle hours of travel on the urban road network (000 hrs) [4] | 2011 | 2097.0 | 2157.2 (2006-7) |
| 5. Melbourne's freeways total daily vehicle hours (000s)[4] | 2011 | 331.0 | 244.7 (2006-07 ) |
| 6. Melbourne's arterial roads total daily vehicle hours (000s)[4] | 2011 | 1766.0 | 1866.8 (2006-07) |
| 7. Freeway travel's share of total vehicle kms of across the urban network [4] | 2001 | (i) 17.3% (ii) 20.3% | 21.9% |
| 8. Total cumulative change in daily vehicle hours (DVH) [5] | 1997-8 to 2006-07 | -162 | 328.7 |

Figure 4.5 – Predicted and actual travel speeds and times, Melbourne

## 4.4 Economic development impacts of transport investments – empirical evidence

A common motivation for investment in transport schemes is to stimulate additional economic growth that would not otherwise have happened without these schemes. Major transport schemes take some time to be planned and delivered, and it can then take time for local and regional economies to respond to the new transport scheme or service. Therefore, long-term evaluation is essential if we are to be able to measure empirically the economic development impacts of new transport investments. A 1998 UK Government report on the links between transport and economic growth stated (p 12) that

*'…we are provided with a strong theoretical expectation that all or part of a successfully achieved transport cost reduction may subsequently be converted into a range of different transport and wider economic impacts. This, in principle, provides for the possibility of improved economic performance. Empirical evidence of the scale and significance of such linkages is, however, weak and disputed. We conclude that the theoretical effects listed can exist in reality, but that none of them is guaranteed. Our studies underline the conclusion that generalisations about the effects of transport on the economy are subject to strong dependence on specific local circumstances and conditions.'*

In line with the above quote, there are still very few empirical studies that have unequivocally been able to demonstrate significant economic development benefits as a result of transport investments, due partly to the lack of long term evaluation data. Banister and Berechman (2000) review case studies of the M25 and a new LRT in Buffalo in the US; neither case study reveals economic growth that can be seen to be additional to that which would have occurred without that transport investment but, rather, that the investment spatially redistributed economic activity according to changing patterns of accessibility.

Lian and Ronnevik (2010) reviewed 102 major road investments completed in Norway between 1993 and 2005. They were unable to establish any relationship between infrastructure investments and employment, income and industrial development, although they did find some evidence that these investments led to some agglomeration effects in regional centres, reducing leakage of economic activity from them to larger Norwegian cities. Some positive labour market effects were also observed by their colleagues Gjerdåker and Øystein Engebretsen (2010) due to regions being strengthened by road investment. In these Norwegian cases the authors were able to use nationally collected data rather than having to collect data from each case themselves.

A review of the links between GDP, GDP growth and transport infrastructure investment in western EU member states (the "old" member states) is another example of long-term evaluation that can be carried out using national data on GDP growth and growth in national road and high speed rail (HSR) network length. The results (from Rye and Scotney, 2011) are shown in the figures below, where each dot represents the situation in each of the "old" 15 member states of the EU, based on nationally available data. These show that there seems to be no close link between transport infrastructure investment and growth in GDP – or, more probably, that GDP and GDP growth are much more affected by other factors and the effect of transport investment is minimal and therefore difficult to pick up.

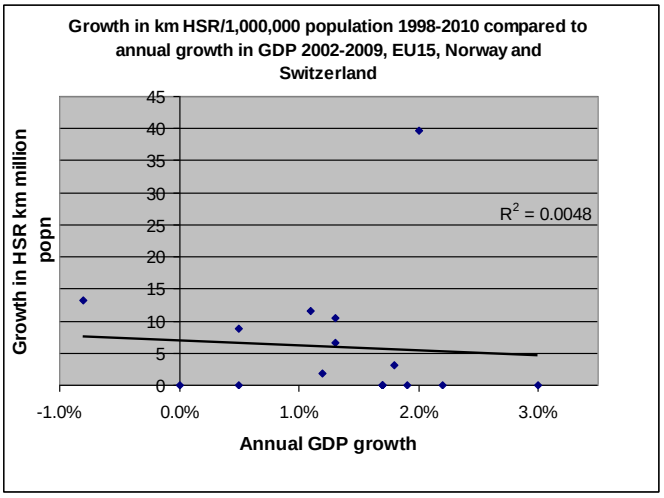


Figure 4.4

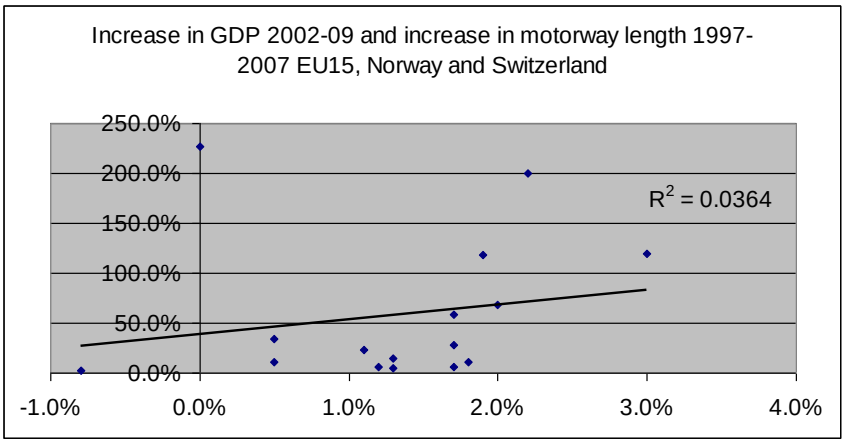


Figure 4.5

## 4.5 Longer term (2-3 years after intervention) evaluation of MM measures

Australia has a relatively long history of mobility management measures focused on households and delivered at the home, known in Australia as “TravelSmart”. Interested households are provided with materials and sometimes incentives (e.g. discount vouchers at sports shops, trial bus tickets and so on) for them to change their travel behaviour. Earlier initiatives in the early 2000s measured the impacts using before and after travel diaries and odometer (car mileage recorder) data. This showed that TravelSmart (TS) households reported greater reductions in their car use than their non-TS counterparts. However, evaluations were done over a short period, typically a year, and so it was not clear whether the changes that TS brought about would be sustained.

In order to deal with this problem, the University of Sydney (Stopher, 2013) was commissioned to carry out a five-year long-term evaluation of households that had participated in TS and others that had not in four urbanised regions in Australia. The number of households participating in the study was small and of course there was a loss of participants from one survey wave to another, but the number of trips recorded was sufficient to be statistically confident of the results. They showed that although all households’ trip making reduced over the period, the TS households started the period with a lower level of car trips than their non-TS counterparts, and that this lower number of car trips remained throughout the five year research period. This is shown in Figure 3.6, below. From this, Stopher concludes that the effects of mobility management in at least these cases were sustained. However, there is a real lack of comparable long-term data from the EU and North America.

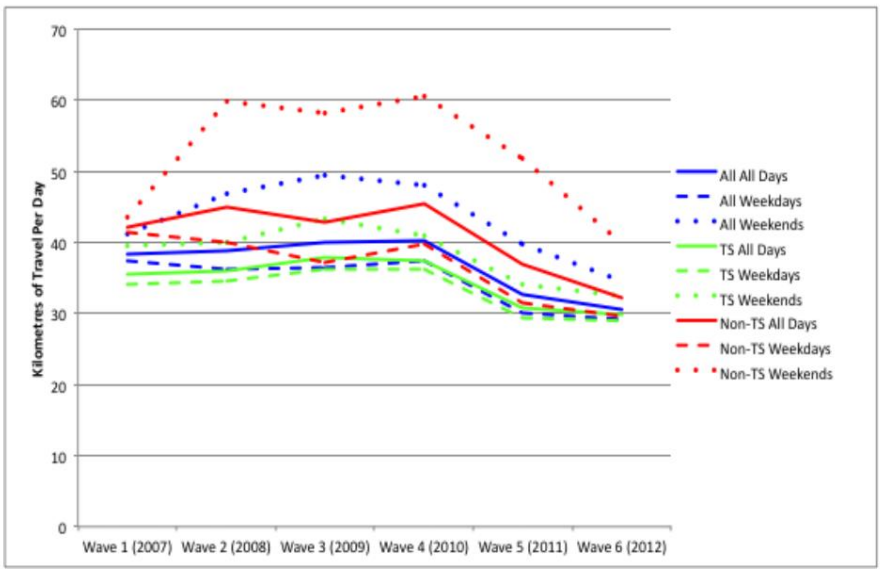


*Figure 3.6 – difference in travel per person per day over 6 survey waves, in TravelSmart (TS) and non-TS households, four Australian urban regions*

# 5 References

Banister D. and Berechman D. (2000) Transport Investment and Economic Development. Psychology Press, Oxford.

City of Edinburgh Council (2007) Third Local Transport Strategy, 2007-2012. City of Edinburgh Council, Edinburgh, UK.

Communite Urbain Strasbourg, 2012. PDU Monitoring Report. Strasbourg, France

Lanzendorf and Busch-Geertsma (2014) The Cycling Boom in German Cities. Transport Policy Journal vol 36 pp 26-33.

Lian J.I and Ronnevik J., Wider economic benefits of major Norwegian road investments, Institute of Transport Economics, Oslo. Paper presented at the World Conference on Transport Research (WCTR) 2010, Lisbon.

Nicolaisen M.S. and Driscoll P.A. (2014) Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review. Transport Reviews 34:4 pp 540-557

Nottingham City Council (2008). Local Transport Plan 2 Monitoring Report. Nottingham City Council, Nottingham, England.

Odgers, J. and Low N. (2010) Travel time savings, transport infrastructure and path dependence, the case of Melbourne, Australia. Paper presented at the World Conference on Transport Research (WCTR) 2010, Lisbon.

Rye, T., and Scotney, D. (2011) Does reducing journey times improve the economy – and, if not, what are the implications for transport assessment? Paper to Scottish Transport Applications and Research Conference, Glasgow.

Stopher, P. (2013) Sustainability of Voluntary Travel Behaviour Change Initiatives: a 5 Year Study. TDM Symposium 2013, Dalian, China.

York City Council (2008). Local Transport Plan 2 Monitoring Report. York City Council, York, England.