

**Risk Analysis of Performance Measure Forecasts in Road Safety  
Engineering**

by

**Craig Alexander Milligan, P.Eng.**

A Thesis Submitted to the Faculty of Graduate Studies of  
The University of Manitoba  
in Partial Fulfilment of the Requirements for the Degree of

**DOCTOR OF PHILOSOPHY**

Department of Civil Engineering  
University of Manitoba  
Winnipeg, Manitoba

Copyright © 2014 by Craig Alexander Milligan

## ABSTRACT

This research contributes to improved risk analysis of performance measure forecasts in road safety engineering by designing and applying a method to characterize uncertainty associated with forecast input data in cases where input uncertainty is not known. The research applies this method to quantify uncertainty in three categories of inputs used in risk analysis of performance measure forecasts in road safety engineering: (1) estimates of pedestrian exposure to collision risk; (2) estimates of vehicular exposure to collision risk; and (3) estimates of engineering economics parameters that assign valuations to mortality risk reductions based on individual willingness to pay. The common methods used in each of these categories are repeated comparisons of input ground truth to input estimations, the use of simulation approaches (e.g. the simulation of short-term counts by sampling permanent count data), and the use of non-parametric techniques to characterize input uncertainty. Some highlights of quantified input uncertainty levels include: (1) when obtaining pedestrian risk exposure estimates at a site in Winnipeg, MB by expanding two-hour short-term counts using the National Bicycle and Pedestrian Documentation Project method, 90% of errors are between -62% and 170%; (2) when obtaining estimates of vehicle exposure to collision risk by expanding two 48-hour counts using the individual permanent counter method for Manitoba highways, 92 % of errors are between -9.5% and 10.8%; and (3) when applying an income-disaggregated transfer function to estimate value of a statistical life for road safety in developing countries, 90% of errors are between -53% and 54%. The results provide further detail on the structure of these input uncertainties. Analytic and computational capabilities in

forecasting and risk analysis have advanced beyond our understanding of corresponding input uncertainty levels; this research closes some of this gap and enables better risk analysis of performance measure forecasts in road safety engineering.

## ACKNOWLEDGMENTS

I thank:

- Dr. Jeannette Montufar, P.Eng., for teaching me many things, for delivering inspiring and dynamic lectures, for serving as my thesis advisor, and for providing formative engineering opportunities;
- Professor Alan Clayton, P.Eng., Dr. Andreas Kopp, and Dr. John Serieux for many formative conversations;
- Dr. Jonathan Regehr, P.Eng., for helpful advice;
- all members of the University of Manitoba Transport Information Group (UMTIG), for contributing to a collegial professional environment; and,
- my family, for love and support.

I acknowledge and appreciate financial assistance provided by the Vanier Canada Graduate Scholarship, the Natural Sciences and Engineering Research Council of Canada (Canada Graduate Scholarship and Michael Smith Foreign Studies Supplement), the World Bank, the Transportation Association of Canada (TAC), Transport Canada, Manitoba Infrastructure and Transportation, the City of Winnipeg, and the University of Manitoba Engineering Graduate Awards Program.

## TABLE OF CONTENTS

1. INTRODUCTION.....	1
1.1. PURPOSE.....	1
1.2. BACKGROUND AND NEED.....	2
1.2.1. Performance Forecasting within Performance Management.....	4
1.2.2. Convergence of Performance Management and Risk Management.....	5
1.2.3. Risk Analysis of Performance Measure Forecasts in Transportation .....	6
1.2.4. Risk Analysis of Performance Measure Forecasts in Road Safety Engineering ...	10
1.3. OBJECTIVES AND SCOPE.....	13
1.4. METHOD.....	17
1.5. THEME AND CONNECTING CONCEPTS.....	18
1.6. THESIS ORGANIZATION.....	19
1.7. THESIS TERMINOLOGY.....	20
2. PERFORMANCE MEASURES AND INPUT UNCERTAINTY FOR PEDESTRIAN CROSSING EXPOSURE ESTIMATES .....	23
2.1. INTRODUCTION.....	24
2.2. EXISTING KNOWLEDGE, PRACTICES and NEEDS.....	26
2.3. METHODOLOGY .....	29
2.3.1. Study Location and Database Development.....	31
2.3.2. Selection of Expansion Methods to Test.....	31
2.3.3. Sample Size & Sampling Method .....	32
2.3.4. Analysis Method .....	34
2.4. RESULTS.....	36
2.4.1. Comparison of Observed Pedestrian Temporal Patterns in Winnipeg with Patterns in Expansion Methods.....	36
2.4.2. Error Distributions Among Expansion Methods .....	43
2.5. DISCUSSION .....	51
2.5.1. Implications of Uncertainty in Performance Measurement .....	51
2.5.2. Relative Influence of Pedestrian Travel Pattern Determinants .....	53
2.5.3. Moving Forward.....	54
2.6. CONCLUSION .....	56
2.7. ACKNOWLEDGEMENTS .....	57
2.8. REFERENCES.....	58
3. ROAD SAFETY PERFORMANCE MEASURES AND AADT UNCERTAINTY FROM SHORT-TERM COUNTS .....	62
3.1. INTRODUCTION.....	63

3.2.	BACKGROUND AND NEED.....	64
3.2.1.	Estimating AADT from Short-Term Counts: Group Factors and Individual Permanent Counter Methods.....	64
3.2.2.	Past Studies on Estimating AADT Uncertainty from Expanded Short-Term Counts.....	67
3.2.3.	Treatment of AADT Uncertainty in Road Safety Performance Measurement.....	69
3.3.	METHOD.....	70
3.4.	RESULTS.....	73
3.4.1.	Summary Characteristics of the Database.....	73
3.4.2.	Distribution of Expansion Transfer Errors.....	74
3.4.3.	Regressions of Transfer Errors on AADT and Distance to Expansion Control Station.....	76
3.5.	DISCUSSION AND CONCLUSION.....	81
3.6.	ACKNOWLEDGEMENTS.....	84
3.7.	REFERENCES.....	84
4.	VALUE OF A STATISTICAL LIFE IN ROAD SAFETY: A BENEFIT-TRANSFER FUNCTION WITH RISK-ANALYSIS GUIDANCE BASED ON DEVELOPING COUNTRY DATA.....	88
4.1.	INTRODUCTION.....	90
4.2.	EXISTING KNOWLEDGE, PRACTICES, AND NEEDS.....	91
4.2.1.	The Transport Safety Problem and the Need for VSL Estimates in Benefit-Cost Analysis.....	92
4.2.2.	Methods to Estimate VSL and their Strengths and Weaknesses.....	94
4.2.3.	Transferring VSL Estimates Between Policy Contexts: Needs, Methods, and Uncertainty.....	98
4.2.4.	VSL and VSI in Developing Countries.....	100
4.3.	METHODOLOGY.....	104
4.3.1.	Data Sources and Methodology for the Development of the Transfer Function..	104
4.3.2.	Methodology to Test for Significant Differences Between Transfer Functions....	114
4.3.3.	Methodology to Quantify and Apply Transfer Function Uncertainty Information.	114
4.4.	RESULTS.....	117
4.4.1.	Transfer function regression results.....	117
4.4.2.	Transfer function errors and the uncertainty analysis.....	121
4.4.3.	Results applied to the transport safety policy context.....	125
4.4.4.	Results for other contexts.....	129
4.5.	CONCLUSIONS.....	130
4.6.	ACKNOWLEDGEMENTS.....	132
4.7.	REFERENCES.....	132

5. CONCLUSION .....	141
5.1. PERFORMANCE MEASUREMENT AND RISK ANALYSIS AS FOUNDATIONS FOR TRANSPORT AGENCY MANAGEMENT .....	141
5.2. CONTRIBUTIONS TO KNOWLEDGE .....	142
5.3. RECOMMENDATIONS FOR FUTURE RESEARCH .....	147
6. REFERENCES.....	148

## LIST OF FIGURES

Figure 1-1: Taxonomy of risk analysis methods .....	8
Figure 2-1: Research Approach: where should we get temporal information to expand short-term counts? .....	30
Figure 2-2: Research approach: how good are expanded short term counts? ...	30
Figure 2-3: Monthly variation in pedestrian volumes .....	37
Figure 2-4: Day-of-week variation in pedestrian volumes .....	37
Figure 2-5: Hourly variation in pedestrian volumes (1: weekday Apr-Sep; 2: single yearly set of factors) .....	38
Figure 2-6: Hourly variation in pedestrian volumes (1: weekday, Oct-Mar; 2: single yearly set of factors) .....	38
Figure 2-7: Hourly variation in pedestrian volumes (1: weekend, Apr-Sep; 2: single yearly set of factors) .....	39
Figure 2-8: Hourly variation in pedestrian volumes (1: weekend Oct-Mar, 2: single yearly set of factors) .....	39
Figure 2-9: Distribution of annual pedestrian crossing exposure estimate errors for each expansion method (Error represents difference between estimates and the reference volume of 1.6 million) .....	45
Figure 2-10: Accuracy of annual pedestrian crossing estimates by expansion method with 90% CI's (based on percentile bootstrap; n = 200, B = 4000) .....	47
Figure 2-11: Precision of annual pedestrian crossing estimates by expansion method with 90% CI's (based on percentile bootstrap; n = 200; B = 4000) .....	47
Figure 3-1: RTE distribution by number and duration of expanded short-term counts .....	75
Figure 3-2: Scatterplots of errors against potential explanatory variables .....	79
Figure 3-3: Residual plots for regressions of error values against explanatory variables .....	81



Figure 4-1: VSL by GDP per capita for estimates passing and failing the quality screening ..... 108

Figure 4-2: Relative transfer errors for the all-income, high-income, and low- and middle-income VSL transfer functions ..... 122

Figure 4-3: VSL by GDP per capita for the transport policy context according to new income-disaggregated benefit-transfer functions ..... 129

## LIST OF TABLES

Table 2-1: Accuracy and precision of annual pedestrian crossing exposure estimates by expansion method with 90% CI' .....	48
Table 2-2: Tests for differences between expansion methods in the accuracy and precision of annual crossing exposure estimates .....	49
Table 3-1: Past studies on vehicle AADT accuracy from expanded short-term counts.....	68
Table 3-2: Database summary .....	73
Table 3-3: Distribution of transfer errors when using individual permanent counter method to expand two 48-hour short-term counts .....	74
Table 3-4: Regression of error values on volume and distance to control station .....	80
Table 4-1: Strengths and weaknesses of the stated and revealed preference methods.....	97
Table 4-2: Overview of surveys and VSL estimate used for developing country transfer function.....	109
Table 4-3: Variables included in regression models .....	113
Table 4-4: Transfer function regression results for the high-, low- and middle-, and all-income functions, with significant differences between function coefficients.....	119
Table 4-5: Percentiles of the relative transfer error distribution with bootstrapped 95% confidence intervals for low- and middle-income, highincome, and all-income transfer functions .....	123

## ACRONYMS, INITIALISMS, AND ABBREVIATIONS

AADT	Annual Average Daily Traffic
AI	All Income (Countries)
AIC	Actual Individual Consumption
ARTE	Absolute Relative Transfer Error
ATE	Absolute Transfer Error
ATR	Automatic Traffic Recorder
B	Number of Resampling Iterations during Bootstrapping
BCA	Benefit Cost Analysis
CI	Confidence Interval
CMF	Collision Modification Factor
FHWA	Federal Highways Administration (United States)
GDP	Gross Domestic Product
GNI	Gross National Income
HI	High Income Countries
IMF	International Monetary Fund
IPC	Individual Permanent Counter Method (for expanding short-term counts)
iRAP	International Road Assessment Programme
LMI	Low Income and Middle Income (Countries)
MAP-21	Moving Ahead for Progress in the 21st Century (U.S. Transportation Funding Legislation)
MARTE	Mean Absolute Relative Transfer Error
MATE	Mean Absolute Transfer Error
MENB	Measurement Error Negative Binomial
MHTIS	Manitoba Highway Traffic Information System
n	Sample size
NBPD	National Bicycle and Pedestrian Documentation Project
NPV	Net Present Value
OECD	Organisation for Economic Cooperation and Development
PDO	Property Damage Only
PPP	Purchasing Power Parity
RED	Roads Economic Decisions Model (World Bank)
RP	Revealed Preference
RTE	Relative Transfer Error
SE	Standard Error
SME	Subject Matter Expert
SP	Stated Preference
SPF	Safety Performance Function
STEAM	Surface Transportation Efficiency Analysis Model (U.S. FHWA)
TAC	Transportation Association of Canada
TE	Transfer Error
TPG	Traffic Pattern Group
VKT	Vehicle Kilometres of Travel

VMR	Variance to Mean Ratio
VOL	Volume
VSI	Value of a Statistical Injury
VSL	Value of a Statistical Life
WHO	World Health Organisation
WTP	Willingness to Pay

## **1. INTRODUCTION**

### **1.1. PURPOSE**

This research presumes that an objective of public investments in road safety is to reduce fatalities and injuries at a reasonable cost and that this objective is often translated into performance measures (such as a benefit-cost ratio) that are forecasted as part of the investment decision process. The performance measure forecasts are uncertain because the inputs and models used to generate the forecasts are uncertain, and this introduces risk to the investment process. Risk is often described as the product of the probability and severity of an undesirable performance outcome. So if a performance measure forecast for a road safety investment reveals an expected benefit-cost ratio of 1.5, a risk analysis of this performance measure forecast might reveal a 30% chance (probability) that the actual benefit cost ratio might be less than 1.0 (severity). Techniques to generate such a risk analysis are well established, but they depend on a robust understanding of uncertainties in the forecast inputs which is often missing.

The purpose of this research is to contribute to improved risk analysis of performance measure forecasts in road safety engineering by designing and applying a method to characterize uncertainty associated with forecast input data. The research applies the method to quantify uncertainty in three categories of inputs used in risk analysis of performance measure forecasts in road safety engineering: (1) estimates of pedestrian exposure to collision risk; (2) estimates of vehicular exposure to collision risk; and (3) estimates of engineering economics

parameters that assign valuations to mortality risk reductions based on individual willingness to pay.

While there are many techniques for risk analysis, this research deals specifically with probabilistic risk analysis, which generates estimates of probability distributions of future performance measure values. These distributions provide not only a most likely performance measure forecast, but also probabilities that the measure will be above or below any given threshold value. The type of probabilistic risk analysis supported by this research generates performance measure distributions by applying performance prediction models within a Monte Carlo simulation. Subjective elicitation from subject-matter experts (Galway, 2007) is a typical method to generate the required input uncertainty information. In this method, the risk analysis practitioner first asks for statements from an expert on the input about the expert's opinion on that input's uncertainty range, and then uses those statements to generate an estimated distribution for that input. An alternative is the use of objective, evidence-based information on input uncertainty acquired from empirical data through measurement, experimentation, and analysis. The main contribution of this research is to support a transition to the use of objective, evidence-based information on input uncertainty from the current use of subjective elicitation from subject-matter experts.

## **1.2. BACKGROUND AND NEED**

The integration of risk and performance management in road safety engineering requires risk-based performance measure forecasting tools and has the potential

to clarify key issues for investment decision-makers. However, a lack of adequate information on uncertainty associated with forecast input data restricts application of these tools to approaches based in subjective elicitation from subject matter experts. Practitioner concerns regarding subjective approaches can exclude risk analysis tools from the decision process entirely.

Within road safety engineering, there are specific needs for evidence-based information on uncertainty associated with pedestrian intersection crossing risk exposure data, vehicle risk exposure data and mortality risk valuation data. These data are key inputs for forecasting the internal rate of return and the net present value of proposed road safety investments, which are performance measures used in the engineering analysis and decision-making processes.

This section establishes the background and need for the proposed research in four sub-sections. Section 1.2.1 contextualizes the use of performance measure forecasts within the broader practice of performance management. Section 1.2.2 explains the current convergence of risk management and performance management in transportation engineering. Section 1.2.3 outlines the state of the practice of risk analysis of performance measure forecasts in transportation engineering. Section 1.2.4 describes the state of the practice of risk analysis of performance measure forecasts in road safety engineering and illustrates the specific knowledge gaps concerning the uncertainties associated with pedestrian crossing risk exposure, vehicle risk exposure and mortality risk valuation data.

### 1.2.1. Performance Forecasting within Performance Management

Transportation performance measures are “a means of summarising the current position and direction and rate of change of progress toward a particular goal or objective” (Marsden, Kelly, & Snell, 2006, p. 22). Transportation agencies that use performance measurement usually do so for these benefits:

- 1) improved decision-making (Karlaftis & Tsamboulas, 2012) (Knoop, Snelder, van Zuylen, & Hoogendoorn, 2012) (Börjesson & Eliasson, 2011) (Henning, Muruvan, Feng, & Dunn, 2011) (Chen & Kasikitwiwat, 2011) (Ramani, Zietsman, Knowles, & Quadrifoglio, 2011) (Cambridge Systematics, Inc. & High Street Consulting Group, 2010) (Hickman & Banister, 2007) (Stradling, Anable, & Carreno, 2007) (Meyer M. , 2005, p. 173) (Moynihan, 2005) (Short & Kopp, 2005) (Francis, Fry, & Humphreys, 2002) (Hauer, Kononov, Allery, & Griffith, 2002) (Yew & Friedman, 2002) (Baird & Stammer Jr., 2000) (Gihring & Greene, 2000, p. 93) (Pratt & Lomax, 1996) (Campbell, 1995) (Gargett & Wallis, 1989) (Lee D. , 1989);
- 2) improved accountability (FHWA, 2012) (Cambridge Systematics, Inc. & High Street Consulting Group, 2010) (Baird & Stammer Jr., 2000) (Bremmer & Bryan, 2008a) (Bremmer & Bryan, 2008b) (Bremmer, Cotton, & Hamilton, 2005) (Gifford & Carlisle, 2004) (Manzetti, 2003) (Yew & Friedman, 2002) (Baird & Stammer Jr., 2000) (Ball & Dunn, 1997) (Putterill & Maani, 1992);  
and,



3) improved focus and effort (Henning, Muruvan, Feng, & Dunn, 2011) (AASHTO, 2011) (Marsden G. , 2009) (Handy, 2008) (Marsden & Bonsall, 2006) (Moynihan, 2006) (Foote, 2004) (Kassoff, 2001) (Abbott, Cantalupo, & Dixon, 1998) (DoT, 1996) (Etmanczyk, 1995).

Studies concerning future values of performance measures, including performance forecast studies, contribute to the overall benefit of improved decision-making by allowing agencies to explore relationships between courses of action and future performance measure values (see, e.g., Hickman and Banister (2007); Börjesson & Eliasson (2011); and Guerre and Evans (2009)). Performance forecasting in transportation engineering dates at least to Manheim's systems analysis approach (1979), which defines the core function of transport systems analysis as the prediction of flows and their associated performance levels. Questions about the reliability of performance forecasts have led to the integration of risk and performance management.

### **1.2.2. Convergence of Performance Management and Risk Management**

There is a natural overlap of performance and risk management. Managing transportation agency performance involves a focus on managing towards outcomes (Abbott, Cantalupo, & Dixon, 1998) which are inherently subject to uncertainty (Marsden & Bonsall, 2006). Managing agency risk involves managing the most important uncertainties faced by an organization (Curtis, D'Angelo, Hallowell, Henkel, & Molenaar, 2012). Risk can be thought of as the product of the severity and probability of an undesirable outcome. To analyze and manage these

risks, a transportation agency requires a categorization and definition of undesirable outcomes. For a coherent management approach, the categorization of undesirable outcomes for risk management should be aligned with the categorization of desired outcomes for performance management.

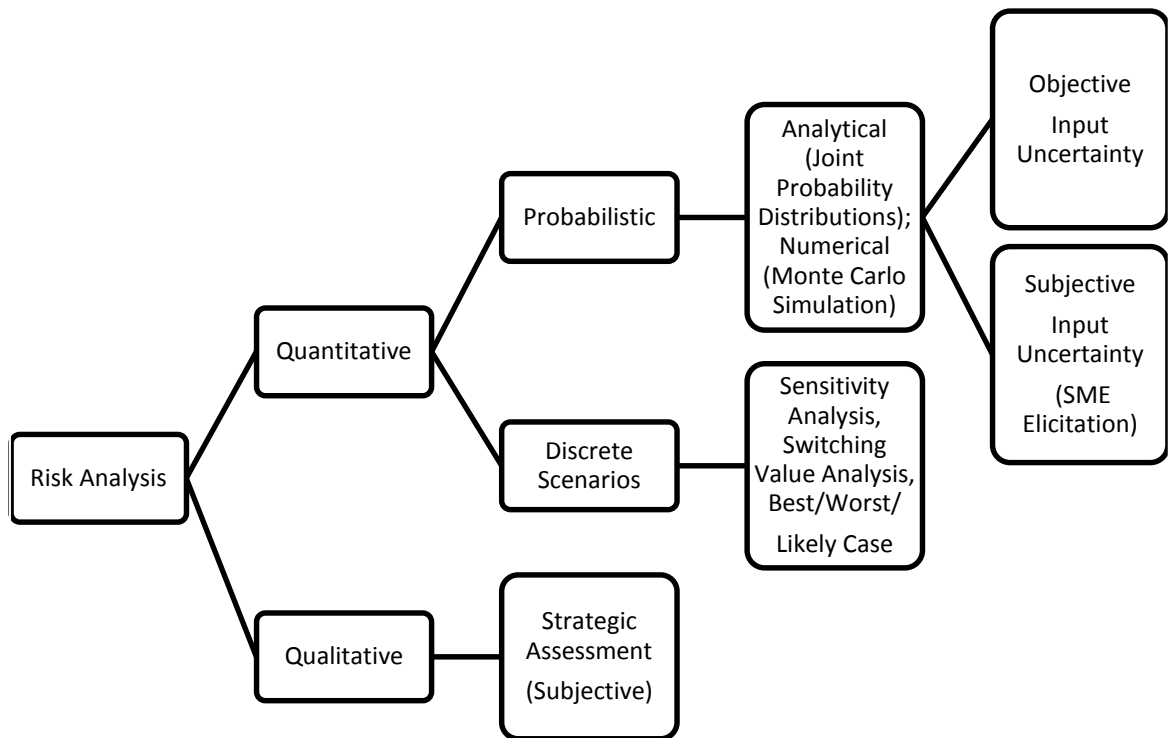
Policies contained in legislation that periodically re-authorizes and sets out conditions for U.S. federal spending on transportation often serve as indicators of major trends with respect to the management of transport investments both in the US and globally (Canning, Hellowell, Hughes, Gatersleben, & Fairhead, 2010) (Mizusawa & McNeil, 2005) (Meyer M. , 2000) (Goldman & Deakin, 2000) (Giuliano, 2007). Policies in the most recent such legislation, *Moving Ahead for Progress in the 21<sup>st</sup> Century* (MAP-21) indicate a convergence of performance management and risk management as key investment decision support tools (US FHWA, 2012). Hauer (1997) argues for the need to know - before a road safety investment is implemented - both the expected impact on reduced collisions (performance prediction) and also the variance around that impact (risk analysis), pointing to the convergence of risk and performance management specifically in the field of road safety.

### **1.2.3. Risk Analysis of Performance Measure Forecasts in Transportation**

Transportation agencies began adapting private sector practices of strategic planning and management in the 1980s (Poister, Margolis, & Zimmerman, 2004) (Meyer, 1988). These practices naturally lead to use of risk analysis techniques because strategic management focuses on internal and external threats to the

attainment of mission-based objectives (Meyer, 1988) (Howard, 1988) (Bishop-Edkins & Nethercut, 1988) (Poister T. , 2004). Since the 1980s, a wide variety of risk analysis techniques have emerged, and a classification scheme for these methods is useful.

Figure 1-1 shows a taxonomy of methods for risk analysis by transportation agencies, with a top-level division between the qualitative methods used in broad strategic planning assessments and the quantitative assessments used in more detailed analyses. Within quantitative approaches, the taxonomy makes a division between those techniques that focus on clarifying a few *discrete* scenarios and those that aim to clarify decisions by providing *probabilistic* information across a continuous range of scenarios. Probabilistic techniques combine information on input uncertainty to estimate output uncertainty, and the taxonomy makes a final distinction based on the source of the input uncertainty information: it can be subjective (elicited from subject matter experts) or objective (based on empirical data). This research supports and enables a transition to the latter.



**Figure 1-1: Taxonomy of risk analysis methods**

Examples of early risk analysis efforts in strategic management that are largely subjective and qualitative in nature are: the environmental scanning process at the Ontario Ministry of Transportation and Communications (Ontario MTC, 1983); the risk assessment procedures at Idaho Department of Transportation (Poister T. , 2004); and the subjective threat assessment matrix at the Massachusetts Department of Public Works (Meyer M. , 1988). On the quantitative side of the taxonomy, examples of discrete scenario analysis can be found in most World Bank project assessment documents for transport projects: key parameters are varied (often by 10 to 20 percent) to illustrate impacts on the forecasted performance measure, which is usually project net present value or internal rate of

return (see, e.g., the assessment for a road safety project in Argentina (World Bank, 2010a)), and the assessment for a highway project in Ningxia, China (World Bank, 2010b)). On the quantitative side of the taxonomy, the alternative to discrete scenario analysis is probabilistic analysis, which determines the probability distribution of the performance forecast based on the probability distributions of the input variables. An example application of analytic techniques to arrive at the outcome probability distribution is given in Ismail and Sayed (2012). Examples of numerical techniques to arrive at the outcome probability distribution are more numerous: major examples are the Monte Carlo simulation models built into the *Roads Economic Decision Model* (Archondo-Callao, 2004) and into the *Surface Transportation Efficiency Analysis Model (STEAM 2.0)* (U.S. FHWA, 2000). The practice of subjectively estimating input uncertainties with judgements from subject matter experts is described in Hertz's seminal work on Monte Carlo-based risk analysis of capital investment in the private sector (Hertz, 1964). Today, the subjective approach either persists or precludes the application of risk analysis altogether because informed practitioners know that the credibility of the risk analysis results depends on the credibility of the probability distributions assumed to represent the input uncertainty ranges, and informed practitioners know that these assumptions often have no empirical basis. Despite these limitations, the demand for increased risk analysis of performance forecasts is likely to increase due to MAP-21 legislation passed in the United States in 2012 that requires increased performance measurement and risk analysis programs on the part of recipients of federal transportation funding.

#### **1.2.4. Risk Analysis of Performance Measure Forecasts in Road Safety Engineering**

One area where risk analysis of performance forecasting has gained traction within road safety engineering specifically is in the field of collision modification factor development. Collision modification factors (CMF) represent a percentage change in collisions expected to accompany the introduction of a road safety countermeasure in a specific context. For example, installation of any type of centre median barrier on a rural multilane divided arterial when there was no median barrier previously is associated with a fatal collision CMF of 0.57; this means that the expected number of collisions after introduction of the measure is 57 percent of the expected number of collisions before the introduction of the measure (U.S. FHWA, 2009). The adjusted standard error for this CMF is .10. Collections of collision modification factors (for example the online CMF clearinghouse (U.S. FHWA, 2013), the *Handbook of Road Safety Measures* (Elvik, Høye, Vaa, & Sørensen, 2009), and the *Highway Safety Manual* (AASHTO, 2010)) all present CMFs with standard error and confidence interval values whenever possible. This information on input uncertainty can aid in risk analysis of road safety performance forecasts.

However, a countermeasure CMF is only one of several forecast inputs subject to uncertainty, and the uncertainty of most other inputs is usually unknown. To illustrate the specific need for the uncertainty information quantified in this research, consider a simple example to forecast the net present value of a road

safety countermeasure introduction where the only benefit considered is reduced fatalities. The net present value is

$$NPV = \sum_{n=0}^y [(1 - CMF_F) * SPF_F(E_n, \{K\}) * VSL - C_n] * (1 + i)^{-n}$$

where

$NPV$  is net present value,

$n$  = the year of analysis,

$y$  = the number of years in the analysis,

$CMF_F$  is the collision modification factor for fatal collisions for the relevant countermeasure, expressing the expected number of collisions *with* the countermeasure as a proportion of the expected number of collisions *without* the countermeasure,

$SPF_F$  is the safety performance function expressing the expected number of collisions for the relevant facility type *without* the countermeasure introduction,

$E_n$  is risk exposure (traffic volume) in year  $n$ ,

$\{K\}$  is a vector of other independent variables in the safety performance function (for example number of lanes, divided status of roadway),

$VSL$  is the value of a statistical life for evaluation purposes,

$i$  is the social discount rate, and

$C_n$  is the cost associated with countermeasure implementation in year  $n$ .

In this formulation, uncertainties in exposure and in the value of statistical life (in addition to uncertainty in the collision modification factor) translate into uncertainty in the performance measure forecast.

A thorough risk analysis for the NPV performance measure in the case illustrated above requires evidence-based uncertainty information pertaining to the risk exposure and value of statistical life inputs. Evidence-based uncertainty information specifically does not exist for:

- 1) pedestrian risk exposure metrics when these metrics are annual volumes derived from expanded short-term counts;
- 2) vehicle risk exposure metrics when these metrics are annual average daily traffic derived from expanded short-term counts using the individual permanent counter method; and,
- 3) value of statistical life estimates for road safety evaluation purposes when these estimates are derived from income-disaggregated meta-analysis benefit transfer functions.

A fundamental reason this uncertainty information does not exist is the lack of a methodology to experimentally develop evidence-based information on the uncertainty associated with these types of inputs. This research is predicated on the specific need for uncertainty information for these three categories of performance measure inputs, and on the more general need for a methodology to



systematically characterize this type of uncertainty in ways practical for subsequent application to risk analysis. This kind of uncertainty information is also not available for several other inputs to safety performance forecasts (for example truck traffic estimates, many safety performance function parameters, many collision modification factors, and valuations for injury collisions). Application of the current methods to these inputs would be a natural extension of this research.

### **1.3. OBJECTIVES AND SCOPE**

The research accomplishes three specific objectives in support of the overall purpose:

**Objective 1:** Quantify uncertainty in estimates of pedestrian risk exposure when these estimates are derived from expanded short-term counts.

**Objective 2:** Quantify uncertainty in estimates of vehicle risk exposure when these estimates are derived from expanded short-term counts using the individual permanent counter method.

**Objective 3:** Quantify uncertainty in estimates of value of a statistical life (VSL) when these estimates are based on between-country transfer function results<sup>1</sup>.

---

<sup>1</sup> A between-country transfer function for VSL gives a VSL estimate for a country based on some variables linked that country's characteristics (for example income levels). The functions are built using data from countries that do have a VSL estimate to facilitate low-cost estimates in countries that do not have one.

The scope for Objective 1 is defined by the pedestrian risk exposure metric considered, the geography of the study, the expansion methods considered, and the temporal scope. The risk exposure metric considered for Objective 1 is annual pedestrian crossing volume at an intersection crosswalk. The question of what constitutes pedestrian exposure is not a simple one, and it has many answers depending on the context and the need for the exposure information. An exposure metric should have some correspondence with the target risk it is being associated with. The research uses annual pedestrian crossing volumes as an exposure metric because it corresponds with the target risk, and the practicality of this metric corresponds with available resources (Milligan, Poapst, & Montufar, 2013). Fitzpatrick and Park (2009) and Zegeer *et al.* (2005, p. 67) use this exposure metric as an input to safety performance measure calculations when researching pedestrian infrastructure design options. The geographic scope for Objective 1 includes one downtown urban intersection pedestrian crossing site in Winnipeg. The major road of the intersection is a six-lane divided arterial street with average weekday daily traffic of about 35,000 vehicles per day and the minor road is a four-lane one-way arterial street with average weekday daily traffic of about 18,000 vehicles per day. The crossing accommodates over 4,000 pedestrians per day. The scope of Objective 1 includes evaluation of uncertainty associated with two expansion methods: expansion of short-term pedestrian counts with National Pedestrian and Bicycle Documentation Project temporal factors (NBPD, 2009) and with local vehicle temporal factors (Hernandez, et al., 2011). The NBPD factors expand short-term counts based on temporal patterns from pedestrians in

other cities. This practice (but not the NBPD factors specifically) is used by Zegeer et al. (2005). The practice of expansion with local vehicle factors is used by Fitzpatrick and Park (2009). The temporal scope of Objective 1 includes one year of data from October 2009 to September 2010.

The scope for Objective 2 is defined by the vehicle risk exposure metric considered, the geography of the study, the expansion method considered, and the temporal scope. The risk exposure metric considered for Objective 2 is annual average daily traffic (AADT) on road segments without permanent count stations. This exposure metric is used together with segment length to calculate annual vehicle-kilometres of travel (VKT), which forms the denominator of collision rates - a widely used road safety performance measure. This exposure metric is also an argument in the safety performance functions which form the basis of the predictive methods in the *Highway Safety Manual* (AASHTO, 2010). The geographic scope of Objective 2 includes only highways in the province of Manitoba at the classification levels *Provincial Trunk Highway* (PTH) and *Public Road* (PR). Traffic volumes on these highways range from a few hundred to 25,000 vehicles per day. The expansion method considered is the individual permanent counter (IPC) method, which the Manitoba Highway Traffic Information System (MHTIS) uses to estimate AADT based on short-term counts. This method links each of about 2000 short-term count sites to one of about 70 permanent counting sites as an individual permanent counter control station. The method expands two 48-hour counts from a short-term count site to an annual volume by using the ratio of the volume for the same 48-hour periods at the control station to the annual volume at the control

station (Poapst, et al., 2012). The temporal scope of Objective 2 includes five years of data.

The scope for Objective 3 is defined by the theoretical approach to establishing value of statistical life (VSL) in road safety, the benefit transfer method, and the geographic scope. The theoretical approach to establishing VSL considered in Objective 3 is the stated preference (SP) willingness to pay (WTP) approach. Other theoretical approaches – the major ones are the revealed preference willingness to pay approach and the human capital approach (Milligan & Kosior, 2013) – were considered but not selected for reasons given in Chapter 4. The stated preference willingness to pay approach establishes VSL as the ratio of willingness to pay for a small change in mortality risk to the size of the risk change offered, when the willingness to pay is elicited from respondents in a contingent valuation survey and taken as the average of responses from a population-representative sample (Milligan & Kosior, 2013). The benefit transfer function considered in Objective 3 is the income-disaggregated meta-analysis based transfer function, which is a new type of benefit transfer function created in this research that will guide future benefit transfer efforts in road safety by the World Bank. The OECD (2012) defines five categories of benefit transfer methods: (1) simple unit value transfer, (2) unit value transfer with income adjustments, (3) unit value transfer for separate age groups, (4) benefit function transfer, and (5) meta-analysis-based transfer function. The transfer function method tested in this research is an extension of the fifth category of transfer methods. This research develops this method extension because the nature of the relationship between VSL and its explanatory variables is found to

vary significantly across income levels, and income-aggregated approaches to meta-analysis, which tend to be dominated by high-income datasets, obscure the nature of the relationships for low and middle income countries. The geographic scope for Objective 3 includes all countries with GDP per capita above \$1268 when expressed in 2005 International Dollars (there are no original values in the meta-analysis dataset from countries with a lower GDP per capita).

The scope definitions have natural implications on the limitations of the results. For geographic, temporal, or methodological contexts outside of the scope covered by the objectives, the actual uncertainty results are not necessarily transferable, although the method to quantify this uncertainty *is* transferable. Furthermore, only a limited number of inputs to safety performance forecasts are considered, and quantifying uncertainty for these inputs could further improve risk analysis capabilities.

#### **1.4. METHOD**

The method for the experimental objectives in this research follows a three-part approach to quantify uncertainty in an input to a performance measure forecast.

The first part of the method involves the identification of a source of ground truth data for an input that is usually estimated.

Once a source of ground truth is identified or created for a class of performance forecast inputs that are usually estimated, the second part of the method involves a test to define the error structure associated with using the usual estimation

methods for these inputs. Application of these usual estimation methods generates an estimated value for the input corresponding to the ground truth value for the input, and the difference between these values is a sample error value for that performance measure forecast input. When this is repeated many times, a large set of sample error values is generated.

The third part of the method for each experiment involves analyzing and interpreting the error structure results for application to evidence-based risk analysis of performance measure forecasts. The primary method for this is the estimation of reference percentiles of the distribution of non-absolute relative errors and (in two of three cases) the estimation of confidence intervals for these reference percentiles. Confidence intervals for the reference percentile values are estimated by applying non-parametric resampling techniques (bootstrapping) to the original set of sample error values, following the guidelines in Mooney and Duval (1993).

The chapters providing research papers that accomplish the individual objectives include more methodological detail for the individual experiments.

## **1.5. THEME AND CONNECTING CONCEPTS**

The three papers presented in this thesis are organized around the central theme of enabling better risk analysis of road safety performance measure forecasts by designing and applying methods to characterize uncertainty in inputs to these forecasts where this uncertainty was previously not well understood. Additional connecting concepts common to the papers are the general approach of

understanding uncertainty by repeated comparisons of input ground truth to input estimations, the use of simulation approaches (e.g., the simulation of short-term counts by sampling permanent count data), and the use of non-parametric techniques to characterize input uncertainty. The non-parametric techniques characterize input uncertainty primarily in terms of reference percentiles of the distribution of input estimate errors without imposing any kind of assumed distributional form. This allows a risk analysis practitioner to apply specific percentiles in a quick sensitivity analysis or to use the whole empirical distribution in a Monte Carlo approach to risk analysis as described in Hertz (1964). In two of the papers, non-parametric resampling techniques are also used to characterize confidence levels in the input uncertainty characterization (e.g., a 5<sup>th</sup> percentile error may be estimated as -62%, with a 90% confidence interval of -53% to -66%, with the confidence interval estimated by the non-parametric percentile bootstrap method (Mooney & Duval, 1993)).

## **1.6. THESIS ORGANIZATION**

The thesis organization follows the “Manuscripts Within a Thesis” approach described in the University of Manitoba Faculty of Graduate studies guidelines (University of Manitoba Faculty of Graduate Studies, 2014). The introduction describes the essence of the research, background literature, the need for the research, objectives and scope, general methods used throughout, the overall theme and connecting concepts linking the individual works, and the contribution to knowledge made.

Following the introduction, three chapters reproduce self-contained research articles, forming the main substance of the thesis and accomplishing the objectives set out in the introduction. Chapter 2 provides a research paper on uncertainty associated with pedestrian crossing exposure estimates. Chapter 3 provides a paper on uncertainty associated with vehicle exposure estimates. Chapter 4 provides a paper on uncertainty associated with value of statistical life estimates. Each paper considers a common input to safety performance measure forecasts, creates a new understanding of uncertainty associated with that input, and thereby allows improved risk analysis of safety performance measure forecasts.

Chapter 5 provides a conclusion to the thesis that summarizes the contributions to knowledge in the context of the road safety, performance measurement, and risk analysis fields. The concluding chapter also makes recommendations for future research.

## **1.7. THESIS TERMINOLOGY**

**Exposure** – “being in a situation which has some risk of involvement in a road traffic accident” (Wolfe, 1982). For example, driving down a road is being in a situation which has some risk, and an exposure metric could be the average daily number of vehicles driving on a road, or crossing a road as a pedestrian is being in a risk situation, and the average daily number of pedestrians could be the metric.

**Performance:** “The carrying out, discharge, or fulfilment of a command, duty, promise, purpose, responsibility, etc.” (Oxford University Press, 2014).



**Performance Measure:** “a means of summarising the current position and direction and rate of change of progress toward a particular goal or objective” (Marsden, Kelly, & Snell, 2006, p. 22).

**Performance Measure Forecast:** “A prediction of the future value of a performance measure based on some input information and forecasting tools or models”

**Performance Measure Forecast Input:** “An input data element, which is itself often estimated and subject to uncertainty, that is used as a variable in a performance measure forecasting tool or model – for example, traffic volume”

**Probabilistic Risk Analysis:** Risk analysis techniques that combine information on input uncertainty to estimate output uncertainty in the form of a probability distribution of possible outcome values, usually with Monte Carlo simulation methods (adapted from Hertz (1964)).

**Risk:** “A product of the probability and severity of an adverse outcome”. (Adapted from the *Level of Risk* concept in ISO 31000 (ISO, 2009, term 3.6.1.8.)

**Risk Analysis:** “clarifying uncertainty for decision-makers.” While there are many techniques for risk analysis, this research deals specifically with *probabilistic risk analysis*, which is further defined above.

**Risk Analysis of Performance Measure Forecast:** “Systematically estimating the likelihoods of various future performance measure values or ranges based on an understanding of forecast inputs and their uncertainty.”

**Road Safety:** “the number of accidents (crashes) by kind and severity, expected to occur on the entity during a specified period” (Hauer E. , 1997, p. 25)

**Uncertainty:** “the quality of being indeterminate as to magnitude or value” (Oxford University Press, 2014).

**Value of Statistical Life:** “the value a given population places *ex ante* on avoiding the death of an unidentified individual” (OECD, 2012, p. 13).

## 2. PERFORMANCE MEASURES AND INPUT UNCERTAINTY FOR PEDESTRIAN CROSSING EXPOSURE ESTIMATES

The material in this chapter is published in (Milligan, Poapst, & Montufar, 2013), and reprinted with permission of co-authors Rob Poapst and Jeannette Montufar. In following the paper, the chapter is self-contained with its own abstract, introduction, conclusion, and references. The author of this thesis had principal responsibility for all aspects of the paper, while co-authors provided advice, reviews, and spreadsheet programming from co-authors.

**ABSTRACT:** Pedestrian safety performance measures often use estimates of annual crossing exposure as inputs – but relatively little information exists on the uncertainty associated with these inputs. This research considers two sources of temporal information for expanding short-term counts: (1) a composite of pedestrian counts from other cities, and (2) local vehicle counts. A database of pedestrian flows from video review covering 12 months and including over 350,000 pedestrian observations provides a known reference annual volume and a set of short-term counts for expansion and testing. The research compares the temporal information sources with observed pedestrian volumes by analyzing the times and magnitudes of volume peaks. The temporal patterns based on local vehicle counts match observed pedestrian patterns more closely than the external composite pedestrian patterns. To quantify exposure estimate uncertainty, the research uses the local vehicle and external composite pedestrian patterns to expand a sample of short term counts to generate a set of 200 annual estimates, and then compares the estimates to the known reference volume. Exposure estimates developed by

expanding counts with local vehicle factors have the lowest errors (mean: -2%; median: -3%, standard deviation: 33%; 90 percent of errors between -53% and 50%). Exposure estimates based on external composite pedestrian patterns have higher errors (mean: 27%; median: 9%; standard deviation: 73%; 90 percent of errors between -62% and 170%). If methods to obtain pedestrian exposure estimates based on short-term counts are improved, more confidence can be placed in safety performance measures that use these estimates as inputs.

## **2.1. INTRODUCTION**

Pedestrian safety analysis often uses exposure information based on expanded short term volume counts. This research considers two fundamental questions related to this practice. First, where should we get temporal information to expand short term counts? Second, how good are the expanded counts?

In addressing the first question, the research compares observed temporal patterns of pedestrian travel at a site to various temporal patterns that could be used for expansion. In addressing the second question, the research compares a known reference annual volume to many estimates based on expanded short term counts in order to analyze the distribution of errors.

The main impetus for this investigation is interest in the potential to use temporal factors from the National Bicycle and Pedestrian Documentation Project (factors based on pedestrian and bicycle counts taken across the United States) to expand short term counts (which are common in many jurisdictions) for use in safety performance measurement and analysis.

This investigation uses extensive video-based data on the volumes of users of a pedestrian facility (an intersection crosswalk in downtown Winnipeg, Canada). For simplicity, we call these users pedestrians, even though the counts include a subset (less than five percent) of users in wheelchairs, strollers, or on bicycles.

Exposure in this research is annual pedestrian crossing volumes at an intersection crosswalk. The question of what constitutes pedestrian exposure is not a simple one and it has many answers depending on the context and the need for the exposure information. An exposure metric should have some correspondence with the target risk it is being associated with. So if the target risk is pedestrian falls, or pedestrians struck by vehicles leaving the road, then time or distance spent walking might be an appropriate exposure metric. If the target risk is pedestrian collisions with vehicles while crossing the street, then number of crossings, number of lanes crossed, time spent crossing streets, number of crossings with pedestrian right-of-way, or number of crossings without pedestrian right-of-way are all possible exposure metrics, among other metrics such as numbers of crossings disaggregated by pedestrian characteristics, lighting conditions, weather, or cross-street traffic characteristics. These metrics would provide a variety of insights for analysis, but some are difficult to record from a practical perspective. This study is concerned with the type of exposure information intended to support the analysis of collisions between vehicles and pedestrians while crossing the street. The research uses annual pedestrian crossing volumes as an exposure metric because it corresponds with the target risk, and the practicality of this metric corresponds with available resources.

## **2.2. EXISTING KNOWLEDGE, PRACTICES AND NEEDS**

Transportation agencies increasingly use performance measurement as a system management tool (Marsden, Kelly, & Nellthorp, 2009). Road safety performance measures are often calculated or analyzed using volume or exposure estimates as inputs. Understanding uncertainty associated with performance measure estimates and the implications of that uncertainty in collision analysis involves understanding uncertainty in performance measure inputs. The ability to make defensible decisions based on a road safety estimate is linked to understanding the uncertainty (bias and variance) of that estimate (Hauer E. , 1997, p. 63). Within data systems for performance measurement and management, there are trade-offs between cost, precision, credibility, and timeliness – users of performance information should be aware of and comfortable with these trade-offs (Zall Kusek & Rist, 2004, p. 86).

For pedestrians, continuous/permanent volume counts are rare (Fitzpatrick & Park, 2009), as is the amount of data on temporal volume variation which could be used to estimate annual exposure based on a short-term count (Aultman-Hall, Lane, & Lambert, 2009). Even though continuous counts are rare, it is known that pedestrian activity in many locations exhibits periodicities with respect to month-of-year, day-of-week, and time-of-day (see, e.g. studies summarized by Schneider et al. (2009)). While information about pedestrian activity at specific times is useful for many applications (e.g. capacity considerations, operational planning, crosswalk warrant studies) information about annual activity is useful for other applications such as developing safety performance functions.

Practitioners are responding to the shortage of pedestrian data through increased installations of new automated continuous counting technologies and through collaborative data efforts. In particular, the National Bicycle and Pedestrian Documentation (NBPD) project has developed a set of publicly available temporal adjustment factors based on a composite of 310 counts in 93 communities throughout the U.S. (NBPD, 2009). The NBPD temporal factors, which are provided for two facility types (multi-use pathways and high-density pedestrian and entertainment areas) and three climate zones (long winter/short summer, moderate, and very hot summer/mild winter), are designed to expand 2-hour pedestrian and bicycle crossing counts to estimates of yearly crossing volumes. The NBPD factors consist of hourly, day-of-week, and monthly factors. There are four sets of hourly factors: summer weekday, summer weekend, winter weekday, and winter weekend. With these four sets, the NBPD method applies more hour-of-day temporal information than a typical expansion effort would (where usually a single set of hourly factors is used (FHWA, 2001)). It is possible that further increasing the number of hourly factor groups used (e.g., with separate hourly factors for Saturdays and Sundays) could further increase accuracy by accounting for different hourly peaking characteristics on these days. This is not addressed in this research, however, since the main impetus is to investigate the NBPD methods.

Several jurisdictional surveys (Hudson et al., 2010, Schneider et al., 2005, Cottrell & Pal, 2003) indicate an increasing focus on continuous automatic pedestrian counts and the temporal information that they can provide for exposure estimates.

Bu et al. (2007) and Ozbay et al. (2010) review costs, capabilities, and technical limitations of technologies available to continuously count pedestrians. These technologies are: infrared counters, piezoelectric mats, laser scanners, computer vision (video), and microwave detection.

Two pedestrian safety studies that resulted in changes to the U.S. Manual on Uniform Traffic Control Devices used techniques to estimate annual crossing exposure by expanding short-term counts. Fitzpatrick and Park (2009) expand two-hour pedestrian counts at each of 123 sites in their study with a two-step approach: (1) two-hour volumes are expanded to daily volumes based on factors from four 24-hour video pedestrian counts taken during the study; and (2) daily volumes are expanded to annual volumes with seasonal vehicle temporal factors from the same city. Zegeer et al. (2005, p. 67) expand 1-hour counts at 2000 sites across the U.S. with temporal factors developed from two sources: (1) 8- to 12-hour pedestrian counts during the study at 22 sites, and (2) 24-hour pedestrian counts conducted 20 years earlier in Seattle.

Previous studies that aim to characterize the accuracy or precision of vehicle traffic volume estimates based on short-term counts follow a four step approach: (1) obtain a reference value through continuous monitoring at a site; (2) sample the continuous monitoring data to create a set of short-term counts; (3) expand short-term counts to create a set of traffic volume estimates using the method being tested; and (4) compare the estimates to the reference value (Jiang et al., 2006) (Yang and Davis, 2002) (Hu, et al., 1998) (Sharma et al., 1996) (Granato, 1998) (Chen S. , 1981). While this four-step approach has been repeatedly applied to

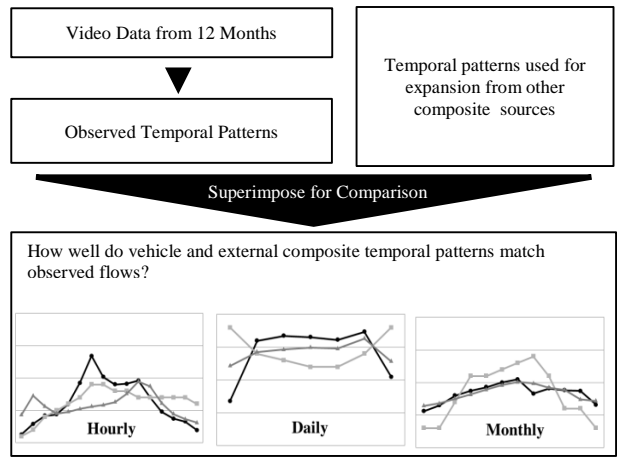


test vehicle volume estimates, it has not been applied to test annual pedestrian volume exposure estimates.

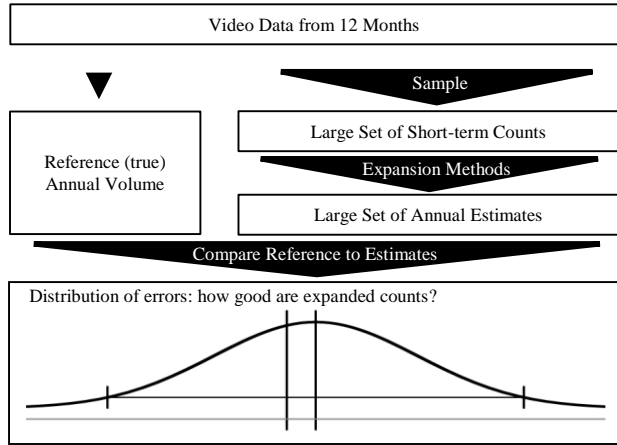
The need for this research is predicated on: (1) the increasing data demands of performance measures; (2) the need to understand uncertainty in performance measures; (3) the need to balance trade-offs between data cost and precision; (4) the lack of pedestrian travel monitoring data; (5) the creation of the NBPD count expansion tools; and (6) the absence of prior studies using a known reference volume to test annual pedestrian crossing exposure estimates based on expanded short-term counts.

### **2.3. METHODOLOGY**

The following questions drive the methodology for this research: (1) where should temporal information for expansion come from (Figure 2-1), and (2) how good are expanded counts (Figure 2-2). Figure 2-1 illustrates that, to answer the first question, the research uses video data of pedestrians at the study location to develop observed hourly, day-of-week, and month-of-year temporal patterns for pedestrian crossing activity. The patterns are in the form of percentages of total volume to facilitate comparison with patterns from other composite sources with different absolute volumes. These graphs then provide a reference point to methodically address differences in periodicities.



**Figure 2-1: Research Approach: where should we get temporal information to expand short-term counts?**



**Figure 2-2: Research approach: how good are expanded short term counts?**

Figure 2-2 illustrates that, to answer the second question, we use a large sample of counts from the continuous video data to develop one reference (close to true) annual volume. The research uses the same continuous video data to sample a large set of short-term pedestrian counts. The research applies the expansion methods under investigation to this sample of short-term counts to generate a large set of annual estimates, generates a corresponding set of errors based on the

difference between these estimates and the reference volume, and analyzes the distribution of these errors to quantify the uncertainty associated with using these methods. The remainder of this section explains the methodology in more detail.

### **2.3.1. Study Location and Database Development**

The research develops a database of pedestrian flows based on a manual review of 84 days of video resulting in 351,000 crossing observations in Winnipeg, Canada. For each of the 84 days counted, a 16-hour count from 06:00 h to 22:00 h, with data in one-hour intervals, is multiplied by 1.05 to obtain a daily reference volume in accordance with the *NBPD* methodology. The intersection of interest is in Winnipeg's central business district and is surrounded by an arena, an exhibition centre, shopping and eating establishments, hotels, bus stops, and office towers. Because of these land use characteristics, the research uses the *NBPD* factors corresponding to the "high-density pedestrian and entertainment area" facility type. The major road of the intersection is a six-lane divided arterial street with a median and with average weekday daily traffic of about 35,000 vehicles per day. The minor road of the intersection is a four-lane one-way arterial street with average weekday daily traffic of about 18,000 vehicles per day. All four corners of the intersection have two curb cuts and ramps, one for each crossing direction. The specific facility considered in this study is the crosswalk traversing the minor street on the north side of the intersection.

### **2.3.2. Selection of Expansion Methods to Test**

This research investigates three expansion methods:

- The *Base Case* expands a short term count without applying any temporal information – using an intentionally naïve assumption that volumes do not fluctuate by time of day, day of week, or month of year. This method is not recommended, but provides a reference to determine the reductions in uncertainty offered by subsequent methods.
- The *NBPD* method uses temporal factors developed from a composite of counts of bicycle and pedestrian volumes in other cities. While the factors can be used to expand just a two-hour count, the *NBPD* recommends expanding three two-hour counts to three annual volume estimates, and then producing a final annual estimate as the average of the three annual estimates. This research follows this recommendation.
- The *Vehicle Factors* method uses temporal factors from a local vehicle traffic pattern group. The factors in this pattern group are based on continuous vehicle traffic monitoring on provincial highways in and around Winnipeg. The research considers the use of local vehicle factors to expand pedestrian counts based on the following: (1) they represent a simple and low cost option; (2) previous research (e.g. Fitzpatrick and Park, 2009) used local vehicle factors; and (3) the similarity in travel patterns of users of different modes within a city may be enough to make the approach viable.

### **2.3.3. Sample Size & Sampling Method**

The first sample size decision concerns how much video review data was required at the site to obtain a reference annual volume. The research obtains the minimum

amount of data to include a volume for at least one day of week for each month of each year in order to conserve resources but still have a sample that is representative of day-of-week and month-of-year variations. This corresponds to a sample size of 84 days. The sampling of the 84 days was random subject to the constraints of taking one week per month while avoiding statutory holidays and days with missing video data. The average of these 84 daily volumes multiplied by 365 represents the reference annual pedestrian crossing exposure estimate. This is not a perfect annual value for comparison - a practical impossibility. Instead, it is called a reference value, because it is a nearly true value used as a reference point for comparison. Because the reference value is itself an estimate, the research evaluates this value based on the sample size and variation within the sample to determine its suitability as a reference point for determining errors. The reference value is 1,605,187 crosswalk users per year, representing an Annual Average Daily Traffic (AADT) of about 4400 users per day. Evaluation using the student's t test statistic estimates a 95 percent confidence interval for this reference value to be within +/- 7 percent of the point estimate. We consider the range of +/- 7 percent to be small enough to be not meaningful from an engineering perspective, and accept the point estimate as a reference for determining errors.

The second sample size decision concerns how many annual estimates to generate by expanding short-term counts. The research uses the percentile bootstrap method for confidence interval estimation. Guidelines in Mooney and Duval (1993) indicate that when using the bootstrap method, parameter estimation improves only slightly for  $B > 1000$ , and that few dispute the quality of parameter

estimation when  $n$  reaches the range of 30 to 50 (where  $n$  is the original sample size and  $B$  is the number of bootstrap resamples). The research conservatively exceeds these guidelines by a factor of four with an original sample size of 200 annual exposure estimates and 4000 bootstrap resamples. The conservative approach is possible because the 84 days required for the AADT formula provide a sufficient database to sample for this many short-term counts. In order to generate the 200 annual exposure estimates, the research requires a sample of two-hour periods for expansion. The research randomly selects two-hour periods for expansion from the 84 days of video counts subject to the constraint of avoiding Monday and Friday counts (a requirement of the NBPD methodology being tested).

#### **2.3.4. Analysis Method**

First, the analysis plots the observed hourly, daily, and monthly temporal patterns from the 84 days of video data together with the patterns built into the expansion methods (the external composite patterns from the NBPD and the local Vehicle Factors). The analysis compares the times and sizes of travel maximums and minimums and methodically addresses differences that are present.

Second, the analysis compares, for the four expansion methods, the distribution of errors that results when many short term counts are expanded and compared to the known reference value for annual exposure.

Each expansion method produces a set of 200 annual pedestrian crossing exposure estimates. For each estimate, the error is  $Error = (Estimate - Reference$

Value) and the percent error is  $Percent\ Error = 100 * (Estimate - Reference\ Value) / Reference\ Value$ .

The analysis plots the distribution of the 200 error values for each expansion method, and identifies the following sample parameters for these values:

- Mean error (indicator of expansion method accuracy);
- Median error (indicator of expansion method accuracy);
- Standard deviation of errors (indicator of expansion method precision); and,
- 90 percent reference interval (indicator of expansion method precision).

The 90 percent reference interval (defined by the 5<sup>th</sup> and 95<sup>th</sup> percentile values of the error distribution) is distinguished from a confidence interval as follows: instead of predicting a range that has a 90 percent likelihood of containing the true distribution mean, the reference interval gives a range of likely error values. This is useful because a practitioner will likely only generate one annual estimate at a site, and the range of likely error values that this one estimate may take is for some purposes as relevant or more relevant than the likely average error of many estimates. The 90 percent reference interval can be thought of as a range of 'not unusual' values.

In order to provide a basis for comparison of error parameters among expansion methods, the research uses the non-parametric technique of bootstrapping (percentile method) to estimate confidence intervals (CIs) for these parameters.

This statistical method is appropriate because parametric techniques require imposed distributional assumptions such as normality (Mooney & Duval, 1993) (Efron & Tibshirani, 1986), and the Kolmogorov-Smirnov test suggests that for each expansion method, it is very unlikely ( $p < 0.01$ ) that the errors are normally distributed. The research also uses bootstrapping (percentile method) to test for statistically significant differences in error parameter values between the methods, using a procedure as described in (Mooney & Duval, 1993) and using  $B = 4000$  resamples. The null hypothesis for each of these tests is shown in Table 2-2 with the test results.

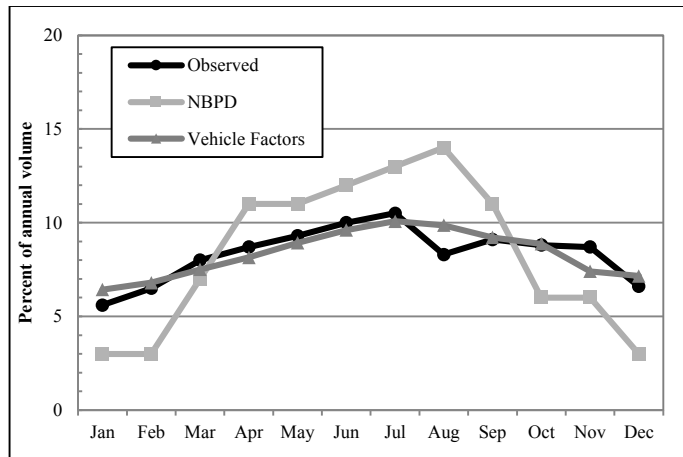
## **2.4. RESULTS**

### **2.4.1. Comparison of Observed Pedestrian Temporal Patterns in Winnipeg with Patterns in Expansion Methods**

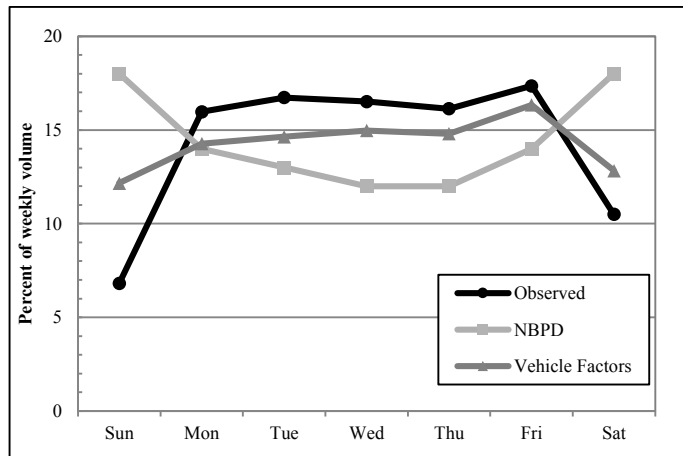
If the periodicities in an expansion method correlate well with actual periodicities at a site, the expansion method is likely to produce good estimates based on short term counts at that site. Figure 2-3 through Figure 2-8 compare observed periodicities at the site to those in the *NBPD* and the *Vehicle Factors* Expansion methods. There are some similarities but also some striking differences. Areas of poor correlation among periodicities can lead to errors in expanded estimates. Figure 2-4 provides an example of how these errors can occur, where the actual observed volumes on Wednesdays are 17 percent of total weekly volume but the *NBPD* expansion factor factors suggest assuming that Wednesday volumes are 12 percent. To expand a Wednesday volume at this site to a weekly volume, it



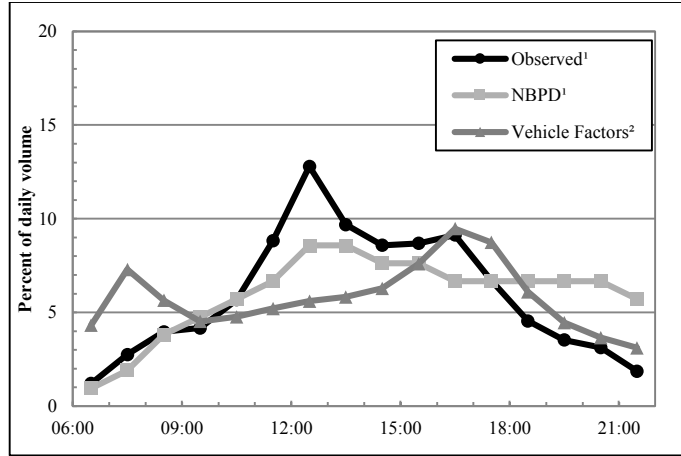
should be divided by 0.17 (based on observed patterns), but using the *NBPD* factors, it would be divided by 0.12, resulting in a weekly estimate 1.42 times what it should be ( $.17/.12$ ). This 42 percent error is one of three errors that is introduced when applying hour-of-day, day-of-week, and month-of-year factors to a two-hour count.



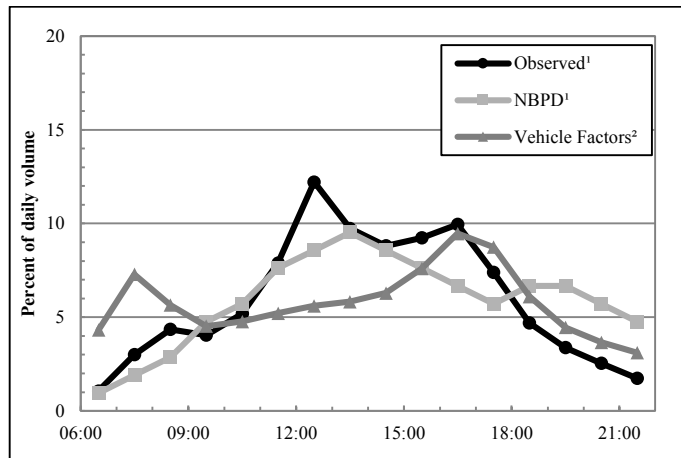
**Figure 2-3: Monthly variation in pedestrian volumes**



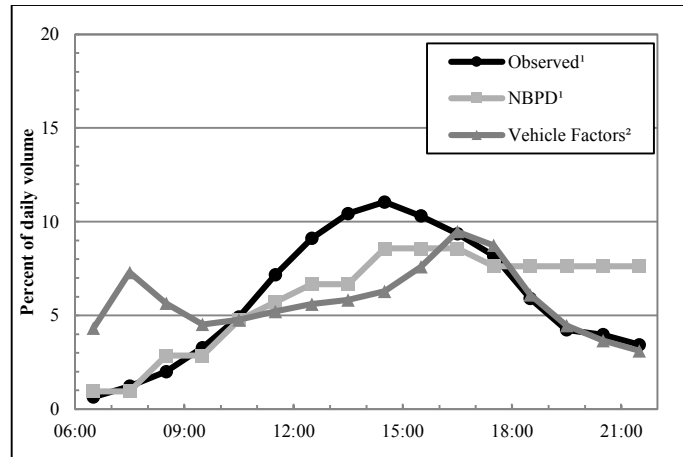
**Figure 2-4: Day-of-week variation in pedestrian volumes**



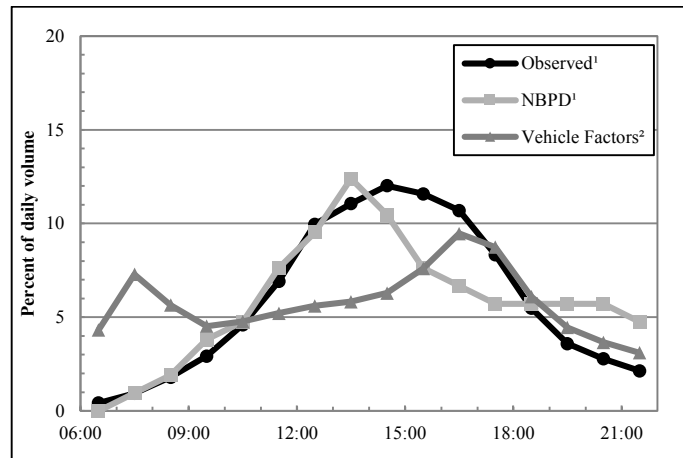
**Figure 2-5: Hourly variation in pedestrian volumes (1: weekday Apr-Sep; 2: single yearly set of factors)**



**Figure 2-6: Hourly variation in pedestrian volumes (1: weekday, Oct-Mar; 2: single yearly set of factors)**



**Figure 2-7: Hourly variation in pedestrian volumes (1: weekend, Apr-Sep; 2: single yearly set of factors)**



**Figure 2-8: Hourly variation in pedestrian volumes (1: weekend Oct-Mar, 2: single yearly set of factors)**

The following discussion focuses on a methodical comparison of the temporal patterns shown in Figure 2-3 through Figure 2-8, based on the locations and magnitudes of maximums and minimums and the impacts of discrepancies in these temporal patterns on expansion errors. It should be noted that the hour-of-day variations shown in Figure 2-5 through Figure 2-8 are separated by type of day (weekday vs weekend) and season (April to September vs October to March),

resulting in four separate graphs. The hourly *Vehicle Factors*, however, as provided by the traffic monitoring program, are not distinguished by time of year or type of day and are the same on all four graphs. If the traffic monitoring program provided separate hourly factors by season and type of day, the match between observed patterns and *Vehicle Factor* patterns in Figure 2-5 through Figure 2-8 might be improved.

Figure 2-3 shows that all three patterns of monthly variations are approximately in phase, exhibiting summer maximums and winter minimums. The amplitude of observed monthly variations closely matches the *Vehicle Factors*, but the amplitude of monthly variations in the *NBPD* factors is more extreme than both of the other patterns. Maximum observed monthly volume is 11 percent of yearly volume, while the *NBPD* maximum is 14 percent. The minimum monthly volume for *NBPD* is 3 percent of yearly volume, which is half of the observed minimum of 6 percent. These discrepancies lead to overestimates of annual exposure when expanding a winter monthly volume with the *NBPD* factors and underestimates of annual exposure when expanding a summer monthly volume. On the other hand, using vehicle factors to expand monthly pedestrian volumes to annual volumes at this site introduces little error.

Figure 2-4 shows that the observed day-of-week variations are approximately in phase with the *Vehicle Factors* (maximum on weekdays and minimum on weekends), but the observed variations have greater amplitude than *Vehicle Factor* variations. The differences in maximums are small: maximum daily observed and *Vehicle Factor* volumes are 17 percent and 16 percent of total

weekly volume, respectively. The differences in minimums are larger: minimum daily observed and *Vehicle Factor* volumes are 7 percent and 12 percent of weekly volume, respectively. These discrepancies lead to slight overestimates of weekly exposure when expanding a weekday daily volume using *Vehicle Factors* and large underestimates when expanding a weekend daily volume.

Figure 2-4 also shows that the *NBPD* factors, with maximum travel on weekend days at 18 percent of weekly volume and minimum travel on weekdays at 12 percent, are completely out of phase with the observed day-of-week pedestrian volume variations. The discrepancies between *NBPD* and observed patterns at the site lead to underestimates of weekly exposure when expanding a weekend daily volume using *NBPD* factors and overestimates when expanding a weekday volume.

Figure 2-5 shows that, during summer months on weekdays, *NBPD* hour-of-day variations in pedestrian volume show a single peak between 12:00 and 13:00 at 9 percent of daily volume. Corresponding observed variations follow a bimodal distribution with peaks occurring between 12:00 and 13:00 at 13 percent of daily volume and between 16:00 and 17:00 at 9 percent of daily volume. The *Vehicle Factors* follow a bimodal distribution with an earlier first peak between 07:00 and 08:00 at 7 percent of daily volume and an second peak between 16:00 and 17:00 at 9 percent. When a two-hour volume is expanded to a daily volume, the discrepancies between observed patterns and those in the expansion methods leads to errors in the daily volume estimate. The size and sign of these errors depends on the time of the count and the expansion method used. Figure 2-6

shows results for hour-of-day variations during weekdays of winter months that are similar to the summer results.

Figure 2-7 and Figure 2-8 show wide peaks in observed weekend pedestrian volumes centred between 14:00 and 15:00 at 11 and 12 percent of daily volume. The bimodally distributed *Vehicle Factors*, which are not specific to weekends, are substantially different from observed hourly variations on weekends. This discrepancy leads to underestimates of daily volumes if a weekend hourly count is expanded using *Vehicle Factors* if the count is taken before 10:00, and overestimates if the count is taken between 10:00 and 16:00. The summer weekend *NBPD* factors show high volumes continuing to 9:00 PM, which is not consistent with observed volumes. This would lead to an underestimate of daily volume if *NBPD* factors were used to expand a summer evening weekend count (which is an unlikely time for an agency to conduct a count, but an important time for pedestrian safety).

The discrepancies among observed temporal patterns and those in the expansion methods result in underestimate and overestimate errors; Section 2.4.2. quantifies and compares these errors. Among all the temporal patterns compared, the best match is between observed volumes and local *Vehicle Factors* in the case of month-of-year variations. This suggests the potential to achieve low annual exposure estimate errors by taking a full week pedestrian count and using only the monthly *Vehicle Factors* for the annual expansion, avoiding the two expansions where there are greater discrepancies in the temporal patterns. The nature of variation in trip-making activity as a function of time can be influenced by a number

of interrelated factors such as trip purpose, land use, feelings of comfort, weather, feelings of security, light levels, opening and closing times of businesses, socio-economic activity patterns, level of service attributes and cost considerations, modal options, and socio-economic attributes of the trip maker. Because these factors vary from place to place (and within places) and from mode to mode (and within modes), discrepancies exist in temporal activity patterns as shown in Figure 2-3 through Figure 2-8. Since using temporal patterns drawn from other modes or cities can be less expensive than obtaining site and mode-specific temporal patterns, the question of whether or not the size of errors resulting from using these methods is acceptable is important.

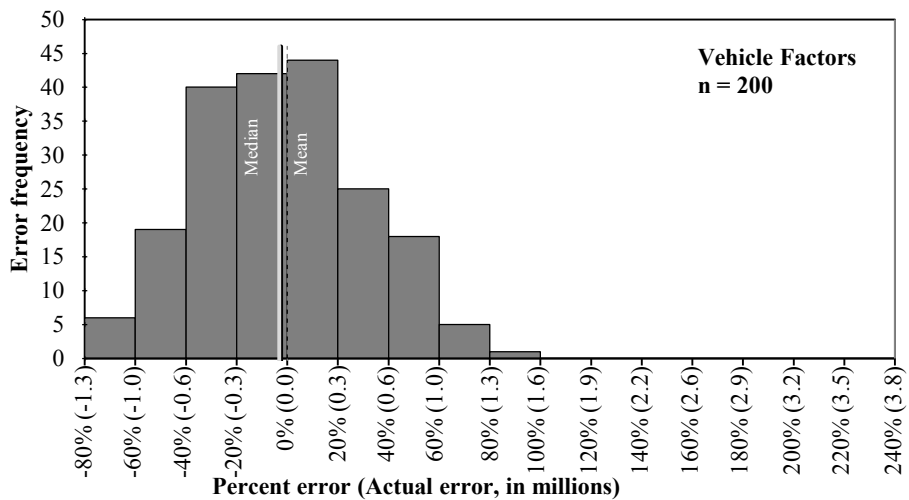
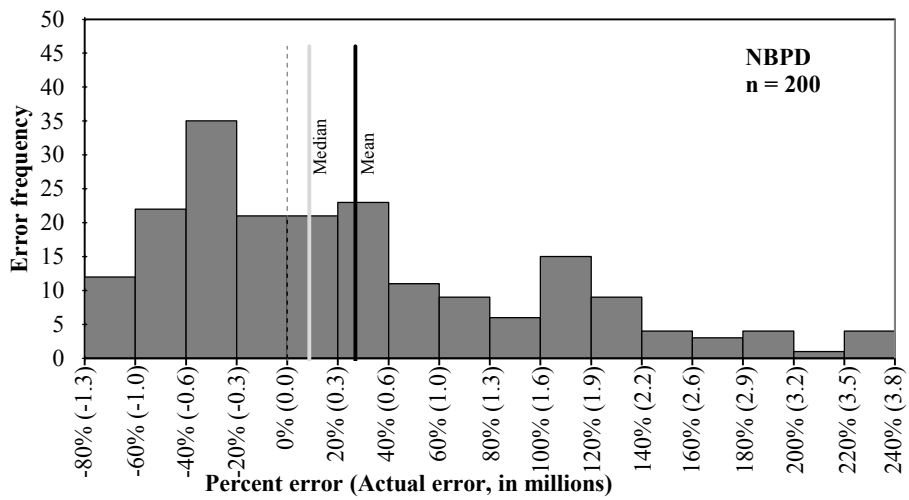
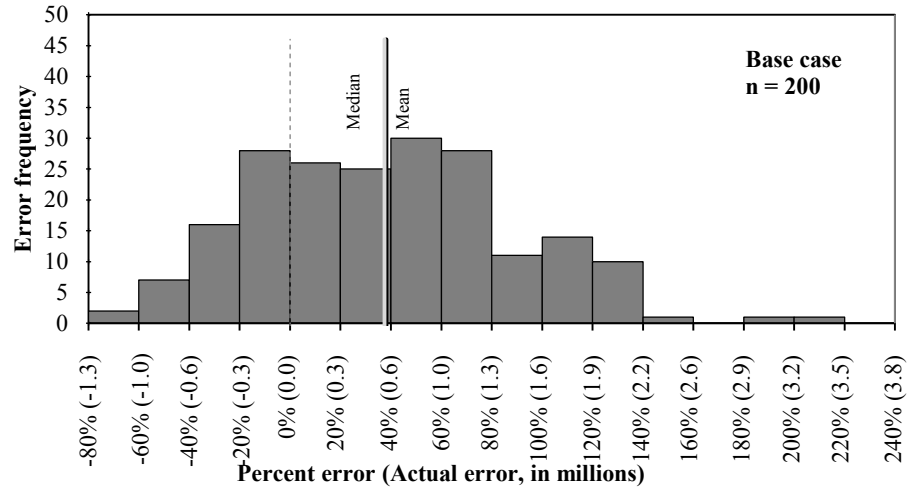
#### **2.4.2. Error Distributions Among Expansion Methods**

This section first presents results of error distributions for various expansion methods, and then presents the results of a test for differences in these parameters between expansion methods. Errors are primarily characterized in terms of percentage deviation from the reference annual volume of just over 1.6 million pedestrians using the crosswalk. For the reader's convenience, in the tables and figures, the error parameters are also presented in terms of actual deviation from the reference annual volumes.

**Figure 2-9** shows a histogram of 200 annual estimate errors for each of the expansion methods. Figure 2-9 shows that the error distributions are generally positively skewed, with long right tail overestimate errors exceeding 200 percent (3.2 million pedestrians) for the *Base Case* and *NBPD* methods. By contrast,

underestimate errors are limited to about -80 percent (-1.3 million pedestrians) for all three methods. The error distribution using the *Vehicle Factors* method is noticeably narrower than the distribution for the other methods, having mean and median errors near zero. **Figure 2-9** suggests a bias towards overestimation of exposure for the *Base Case* and *Vehicle Factors* methods.

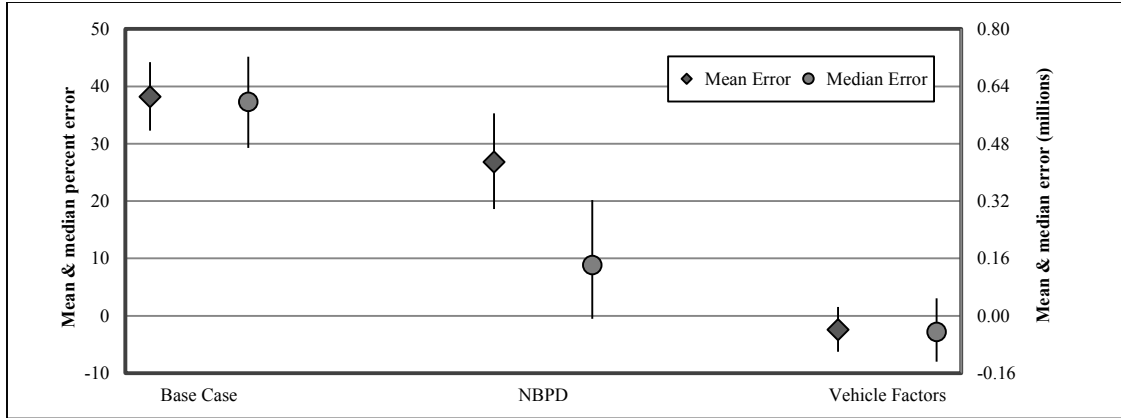




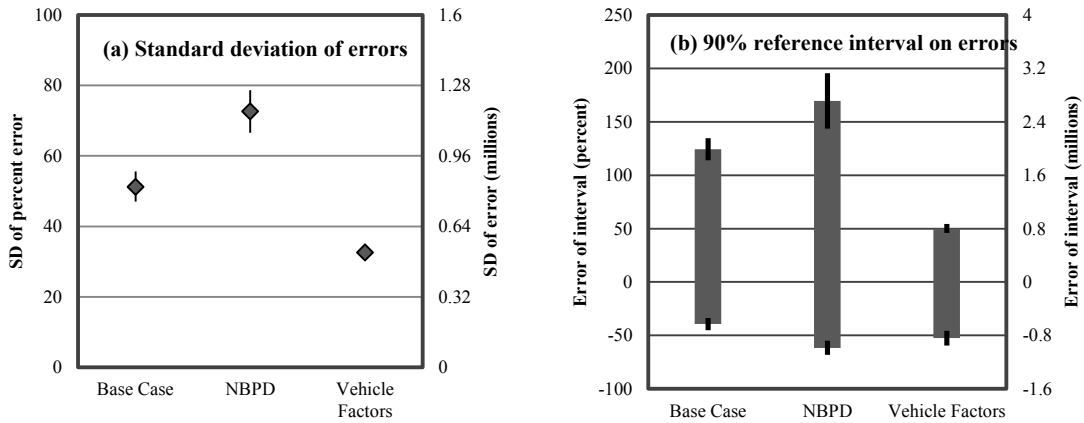
**Figure 2-9: Distribution of annual pedestrian crossing exposure estimate errors for each expansion method.** Note: (Error represents difference between estimates and the reference volume of 1.6 million)

Figure 2-10, and Figure 2-11, and

Table 2-1 highlight several parameters of these error distributions and estimates of 90 percent confidence intervals for these parameters. In terms of accuracy, the mean and median errors for the *Base Case* are both around 40 percent (.6 million pedestrians) revealing a bias towards overestimation significant at the 90 percent confidence level when no temporal information is applied during expansion. For the *NBPD*, mean error is about 30 percent (.48 million pedestrians), which is significantly different from zero at the 90 percent confidence level, while the median error is about 10 percent (.16 million pedestrians), which is not significantly different from zero at the 90 percent confidence level. This reveals a bias towards overestimation using the *NBPD* method. The *NBPD* method has a tendency at this site to produce some estimates with very large errors. For the estimates based on *Vehicle Factors*, mean and median errors are nearly zero and not significantly different from zero, revealing no bias in the method. In terms of precision, the standard deviation of errors is about 50 percent (.8 million pedestrians) for the *Base Case*, 70 percent (1.2 million pedestrians) for the *NBPD* and about 30 percent (.5 million pedestrians) for the *Vehicle Factors*. The 90 percent reference intervals show ranges of normal errors; one in ten errors is more extreme than this range. For the *Base Case*, the range is about -40 percent to 120 percent, for the *NBPD* the range is about -60 to 170 percent, and for the *Vehicle Factors*, the range is about -50 percent to 50 percent.



**Figure 2-10: Accuracy of annual pedestrian crossing estimates by expansion method with 90% CI's (based on percentile bootstrap; n = 200, B = 4000)**



**Figure 2-11: Precision of annual pedestrian crossing estimates by expansion method with 90% CI's (based on percentile bootstrap; n = 200; B = 4000)**

**Table 2-1: Accuracy and precision of annual pedestrian crossing exposure estimates by expansion method with 90% CI's**

Error Parameters	Expansion Method		
	Base Case	NBPD	Vehicle Factors
<b><i>Accuracy Parameters</i></b>			
Mean error, percent	38% (32% to 44%)	27% (19% to 35%)	-2% (-6% to 2%)
Mean error, millions	0.6 (0.5 to 0.7)	0.4 (0.3 to 0.6)	0.0 (-0.1 to 0.0)
Median error, percent	37% (29% to 45%)	9% (-1% to 20%)	-3% (-8% to 3%)
Median error, millions	0.6 (0.5 to 0.7)	0.1 (-0.0 to 0.3)	-0.1 (-0.1 to 0.0)
<b><i>Precision Parameters</i></b>			
SD of errors, percent	51% (47% to 46%)	73% (67% to 79%)	33% (30% to 35%)
SD of errors, millions	0.8 (0.8 to 0.7)	1.2 (1.1 to 1.3)	0.5 (0.5 to 0.6)
5th percentile error, percent	-39% (-33% to -44%)	-62% (-53% to -66%)	-53% (-44% to -58%)
5th percentile error, millions	-0.6 (-0.5 to -0.7)	-1.0 (-0.8 to -1.1)	-0.8 (-0.7 to -0.9)
95th percentile error, percent	124% (114% to 135%)	170% (139% to 191%)	50% (46% to 54%)
95th percentile error, millions	2.0 (1.8 to 2.2)	2.7 (2.2 to 3.1)	0.8 (0.7 to 0.9)

Note: For each method, parameter and CI estimates based on percentile bootstrap with n = 200 and B = 4000.

**Table 2-2: Tests for differences between expansion methods in the accuracy and precision of annual crossing exposure estimates**

Percent Error Parameter	Symbol	Null Hypothesis	Result (percent)	Result (millions)	Two-tailed p-value
<u>Base Case – NBPD</u>					
Mean Error	$\mu$	$H_{0,a}: (\mu_{BASE} - \mu_{NBPD}) = 0$	11%	0.2	0.07
Median Error	X.5	$H_{0,b}: (X_{.5,BASE} - X_{.5,NBPD}) = 0$	28%	0.4	<0.0005
Standard Deviation of Errors	$\sigma$	$H_{0,c}: (\sigma_{BASE} - \sigma_{NBPD}) = 0$	-21%	-0.3	<0.0005
90% Ref. Interval Lower Bound	X.05	$H_{0,e}: (X_{.05,BASE} - X_{.05,NBPD}) = 0$	22%	0.4	0.0015
90% Ref. Interval Upper Bound	X.95	$H_{0,d}: (X_{.95,BASE} - X_{.95,NBPD}) = 0$	-45%	-0.7	0.0035
<u>Base Case – Vehicle</u>					
Mean Error	$\mu$	$H_{0,f}: (\mu_{BASE} - \mu_{VEH}) = 0$	41%	0.7	<0.0005
Median Error	X.5	$H_{0,g}: (X_{.5,BASE} - X_{.5,VEH}) = 0$	40%	0.6	<0.0005
Standard Deviation of Errors	$\sigma$	$H_{0,h}: (\sigma_{BASE} - \sigma_{VEH}) = 0$	19%	0.3	<0.0005
90% Ref. Interval Lower Bound	X.05	$H_{0,j}: (X_{.05,BASE} - X_{.05,VEH}) = 0$	13%	0.2	0.01601
90% Ref. Interval Upper Bound	X.95	$H_{0,i}: (X_{.95,BASE} - X_{.95,VEH}) = 0$	74%	1.2	<0.0005
<u>NBPD – Vehicle</u>					
Mean Error	$\mu$	$H_{0,k}: (\mu_{NBPD} - \mu_{VEH}) = 0$	29%	0.5	<0.0005
Median Error	X.5	$H_{0,l}: (X_{.5,NBPD} - X_{.5,VEH}) = 0$	12%	0.2	0.0775
Standard Deviation of Errors	$\sigma$	$H_{0,m}: (\sigma_{NBPD} - \sigma_{VEH}) = 0$	40%	0.6	<0.0005
90% Ref. Interval Lower Bound	X.05	$H_{0,o}: (X_{.05,NBPD} - X_{.05,VEH}) = 0$	-9%	-0.1	0.1205
90% Ref. Interval Upper Bound	X.95	$H_{0,n}: (X_{.95,NBPD} - X_{.95,VEH}) = 0$	120%	1.9	<0.0005

Notes: Results based on percentile bootstrap with n=200 and B=4000; result represents mean of 4000 differences, two-tailed p-value based on the percentile at which the distribution of 4000 differences becomes more extreme than zero; p<0.0005 indicates that none of the 4000 differences was more extreme than zero.

Table 2-2 shows the results of a test for differences in error parameters between the expansion methods. In terms of accuracy, the results indicate that the NBPD

performs better than the *Base Case*, especially when considering median error which is lower for the *NBPD* by almost 30 percent with  $p < 0.0005$ . In terms of precision, however, the *Base Case* performed better than the *NBPD*, with the lower and upper bounds of the 90 percent reference interval for the *NBPD* more extreme than the *Base Case* by 22 percent and 45 percent, respectively. Using external composite pedestrian factors improves overall accuracy but also introduces the possibility of some very large errors vis-à-vis the *Base Case* alternative of using no temporal information to expand the count.

The results show that the *Vehicle Factors* also outperforms the *Base Case* in terms of accuracy with lower mean and median errors by 41 and 40 percent with  $p < 0.0005$ . In terms of precision, the *Vehicle Factors* produces a standard deviation of errors about 20 percent lower than that produced by the *Base Case*, with  $p < 0.0005$ . The 5<sup>th</sup> percentile error for the *Vehicle Factors* is more extreme than the same parameter for the *Base Case* by 13 percent with  $p = 0.016$ , but the 95<sup>th</sup> percentile error for *Vehicle Factors* is less extreme than the *Base Case* by 74 percent with  $p < 0.0005$ .

The test comparing *Vehicle Factors* to the *NBPD* indicates that the *Vehicle Factors* produce a substantially lower mean error (difference of 29 percent,  $p < 0.0005$ ), but only a moderately lower median error (difference of 12 percent,  $p = 0.077$ ). Compared to the *NBPD*, the *Vehicle Factors* also produce a substantially lower standard deviation of errors (difference of 40 percent,  $p < 0.0005$ ), and a substantially lower 95<sup>th</sup> percentile error (difference of 120 percent,  $p < 0.0005$ ).

## **2.5. DISCUSSION**

This discussion first explores, through illustrative examples, the safety performance measurement implications of exposure uncertainty revealed in the test. Second, it examines what the results may mean with respect to the relative influence of travel mode, geography, climate, and land use on pedestrian travel patterns. Finally, it presents some considerations for moving forward, given the characteristics and implications of exposure estimate uncertainty revealed in the test.

### **2.5.1. Implications of Uncertainty in Performance Measurement**

While *Vehicle Factors* performed better than *NBPD* (in terms of accuracy and precision), and *NBPD* performed better than the *Base Case* (in terms of accuracy), annual estimates of pedestrian crossing exposure based on all methods tested are associated with high uncertainty. When exposure estimates based on these methods are used to generate safety performance measures or analyses, the uncertainty is transferred to the results. This investigation demonstrates uncertainty when the four methods tested were applied in the Winnipeg, Manitoba context – applications in other contexts or with variations in methodology may result in lower or higher uncertainty. The single site is a limitation of the study, and future research can clarify the ranges of uncertainty at other sites.

The 90% reference intervals, seen in Figure 2-11, show a range of ‘not unusual’ error values. In the best case (expansion based on vehicle factors), this reference



interval is about -50 to 50 percent. This reference interval of -50 to 50 percent is useful for thinking about the implications of using annual exposure estimates as an input for safety performance analysis, when that estimate is generated by expanding a short term count with a method similar to one tested in this study (keeping in mind that estimates in other contexts may have different uncertainty characteristics):

- It would not be unusual for the denominator of a collision rate to be off by half.
- It would not be unusual for the argument of a safety performance function (SPF) to be off by half.
- Collision modification factors (CMFs), which may be based on comparing collision rates or SPFs of various entities or of the same entity before and after a treatment, may need to be re-thought in light of the following questions: (1) is a treatment CMF so large that even with possible 50% errors in the inputs, the presence and direction of a treatment effect is still known with any confidence? and (2) to what extent does the sample size of a CMF study mitigate impacts of this uncertainty by allowing overestimates of exposure at some sites to balance underestimates of exposure at other sites?
- A wide variety of economic analyses attempt to support decisions by comparing the cost of installing a treatment to a pecuniary estimate of the benefit of reduced accidents. These calculations often involve projecting

future exposure using an estimate of current exposure together with growth assumptions. Uncertainty in the initial exposure estimate could translate into corresponding uncertainty in the results of the economic analysis. When this uncertainty is quantified, it can be accounted for in risk-based analysis and forecasting.

### **2.5.2. Relative Influence of Pedestrian Travel Pattern Determinants**

The fact that the local *Vehicle Factors* method outperformed the *NBPD* method in both accuracy and precision suggests that pedestrian travel patterns may have more in common with local vehicle travel patterns than they have in common with external pedestrian travel patterns – even when the external patterns are taken from composites of locations with broadly similar land use and climate characteristics. It may also mean that the facility/land use categorizations contained in the *NBPD* are still too coarse for widespread application. For example, the fact that the actual day-of-week variations at the site were completely out of phase with the *NBPD* factors means that the overall facility/land use category “high-density pedestrian and entertainment area” does not adequately distinguish between areas that are and are not influenced by high work-related weekday traffic in addition to shopping, entertainment, recreation, and other discretionary traffic that tends to concentrate on weekdays. Since the *NBPD* is an ongoing project, this may improve in the future. The group working on the *NBPD* plans to update their factors as they continue receiving counts from collaborating jurisdictions.

### **2.5.3. Moving Forward**

Based on our results, several considerations emerge for moving forward. First, performance measures and predictions of performance measures using pedestrian exposure estimates as inputs should be interpreted with caution. When uncertainty characteristics of a performance measure estimate are known, there are cases when that uncertainty information can and should be incorporated to integrate risk-based and performance-based transportation planning. A program of road safety investments may be based on – and justified through - predicted safety improvement benefits which may or may not be realized. This uncertainty poses some degree of risk to a transportation agency's strategic mission. For example, an agency may have a strategic goal of reducing pedestrian crossing fatalities by 30 percent. If the agency implements improvements expected (by point estimates of present and future performance measures) to achieve this goal, there is a good chance they will not achieve it. With uncertainty information, the agency can perform a stochastic analysis using Monte Carlo simulations to design a program with an estimated likelihood of meeting their strategic target. The cost of the program would likely go up with the predicted likelihood of meeting the target, creating a trade-off scenario. Integrating risk-based and performance-based management means making an informed decision about this tradeoff.

Second, for now, it may be better to use local vehicle temporal patterns rather than external pedestrian patterns if expanding a short-term pedestrian count to be used in safety performance measurement and analysis. The word *may* should be stressed – on the one hand, these results are based on one site in the Winnipeg,

Manitoba context, and repeated studies elsewhere could confirm or call into question this result, but on the other hand these results are based on a very large sample of observed pedestrians (84 days and over 350,000 observations).

Third, increased deployment of continuous count technologies, together with increased research on the determinants of pedestrian travel patterns, could eventually lead to better exposure estimates to be used as inputs for safety performance measure calculation and analysis. Schneider et al. (2012) and Miranda-Moreno and Fernandes (2012) are examples of research using increased installations of continuous pedestrian counters. One specific study that could be undertaken in this regard is to install continuous counters in cities where local vehicle factors are available. The present paper found very low median and mean errors with the vehicle factors method. A multiple city study could determine the consistency of this result across jurisdictions and facility types. The present paper also found a very good match between the local *seasonal* vehicle patterns and observed pedestrian flows (a better match than *hourly* and *day-of-week* patterns). If this is consistent in a multiple city vehicle factors study, then a practitioner or researcher seeking pedestrian crossing exposure information could employ the strategy of using automatic counters for one week at a time and using available local vehicle factors to expand these counts to good yearly estimates. This strategy has the potential to provide about 50 times higher return on investment in terms of exposure data per automatic pedestrian counter compared to the strategy of leaving a counter at one site for an entire year. A second study could investigate the impact of using a finer resolution facility typology to create external composite

pedestrian factors. The present study suggests that “high-density pedestrian and entertainment area” is not a discrete enough facility type for one temporal pattern group. A study of a finer resolution facility typology could define facility types, for example, according to relevant characteristics identified in (Schneider, Henry, Mitman, Stonehill, & Koehler, 2012), and then create a multiple city study group with representation of each facility type in each city. For each facility type, the temporal patterns in each city can be compared to the composite of temporal patterns from same facility type from the other cities. The NBPD is an ongoing collaborative project and could be improved if the results from such a study were positive.

## **2.6. CONCLUSION**

This paper investigates the accuracy and precision of pedestrian exposure estimates using four methods to expand short term counts. A database of pedestrian flows developed from video review covering 12 months provides a reference annual volume and a sample set of short-term counts for expansion and testing. The method applying local vehicle temporal factors yields the best precision and accuracy results. The method applying no temporal factors yields the worst accuracy, while the methods applying temporal patterns based on a composite of external pedestrian counts yields the worst precision. Pedestrian travel patterns may have more in common with local vehicle travel patterns than external pedestrian travel patterns.

While the application of vehicle factors produced the best results, the range of observed errors is still quite high: over- or under-estimates of exposure by about 50 percent were not uncommon in the data. A main reason for the high errors is the difference between the periodicities at the site and the periodicities assumed in the temporal patterns used for expansion. A source of temporal information can be used for good expansions only if the periodicities in the source correlate well with the periodicities in activity at the site. When exposure estimates with associated uncertainty are used as inputs for safety performance measure calculation and analysis, there are wide-ranging implications: the use of rates, safety performance functions, collision modification factors, and economic impact assessments is affected. Given the significant degree of uncertainty associated with annual estimates of pedestrian crossing volumes based on expanded short-term counts, future work to install continuous count technologies and investigate the determinants of travel patterns could be useful.

## **2.7. ACKNOWLEDGEMENTS**

This work was supported by funding from the Natural Sciences and Engineering Research Council of Canada and the City of Winnipeg Department of Public Works. The authors also gratefully acknowledge helpful statistics-related comments from Dr. Depeng Jiang, the assistance of those who conducted the video review, and the helpful suggestions by anonymous reviewers.

## 2.8. REFERENCES

Aultman-Hall, L., Lane, D. & Lambert, R. R., 2009. Assessing Impact of Weather and Season on Pedestrian Traffic Volumes. *Transportation Research Record* 2140, 35-43.

Bu, F., Greene-Roesel, R., Diogenes, M. C. & Ragland, D., 2007. Estimating Pedestrian Accident Exposure: Automated Pedestrian Counting Devices Report. UC Berkeley Safe Transportation Research & Education Center, Berkeley, CA.

Chen, S., 1981. Improved Methods for State-Wide Vehicle Counting. Purdue University Joint Highway Research Project, West Lafayette, Indiana.

Cottrell, W. & Pal, D., 2003. Evaluation of Pedestrian Data Needs and Collection Efforts. *Transportation Research Record, Journal of the Transportation Research Board* 1828, 12-19.

Effron, B. & Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, Volume 1, pp. 54-75.

FHWA, 2001. *Traffic Monitoring Guide*, Washington, DC: U.S. Department of Transportation Federal Highway Administration.

Fitzpatrick, K. & Park, E. S., 2009. Safety Effectiveness of HAWK Pedestrian Treatment. *Transportation Research Record, Journal of the Transportation Research Board* 2140, 214-223.

Granato, S., 1998. The Impact of Factoring Traffic Counts for Daily and Monthly Variation in Reducing Sample Counting Error. Iowa State University Center for Transportation Research and Education, 122-125, Ames, Iowa,

Hauer, E., 1997. Observational Before-After Studies in Road Safety. Elsevier Science Ltd, Oxford.

Hudson, J., Qu, T.-B. & Turner, S., 2010. Forecasting Bicycle and Pedestrian Usage and Research Data Collection Equipment. Texas Transportation Institute, College Station, TX.

Hu, P. S., Wright, T. & Esteve, T., 1998. Traffic Count Estimates for Short-Term Traffic Monitoring: Simulation Study. Transportation Research Record 1625, 26-34.

Jiang, Z., McCord, M. R. & Goel, P. K., 2006. Improved AADT Estimation by Combining Information in Image- and Ground-Based Traffic Data. Journal of Transportation Engineering, 523-530.

Marsden, G., Kelly, C. & Nellthorp, J., 2009. The likely impacts of target setting and performance rewards in local transport. Transport Policy (16) 59-97.

Miranda-Moreno, L. F. & Fernandes, D., 2012. Pedestrian Injury Risk at Signalized Intersections: Exposure Measures and Geometric Designs. In: Proceedings of the Transportation Research Board 91<sup>st</sup> Annual Meeting, Washington DC.



Mooney, C. Z. & Duval, R. D., 1993. Bootstrapping: A nonparametric approach to statistical inference (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-095). Sage, Newbury Park, CA.

NBPD, 2009. *Extrapolation Workbook, Extrapolation Detailed Explanation, and NBPD Facts and FAQs*. [Online] Available at: <http://bikepeddocumentation.org/> [Accessed 30 11 2011].

Ozbay, K. et al., 2010. Automatic Pedestrian Counter. New Jersey Department of Transportation, Trenton, New Jersey.

Schneider, R. J., Arnold, L. S. & Ragland, D. R., 2009. Methodology for Counting Pedestrians at Intersections: Use of Automated Counters to Extrapolate Weekly Volumes from Short Manual Counts. *Transportation Research Record: Journal of the Transportation Research Board* 2140, 1-12.

Schneider, R. J. et al., 2012. Development and Application of San Francisco Pedestrian Intersection Volume Model. In: *Proceedings of the Transportation Research Board 91<sup>st</sup> Annual Meeting*, Washington, DC.

Schneider, R., Patton, R., Toole, J. & Raborn, C., 2005. Pedestrian and Bicycle Data Collection in United States Communities: Quantifying Use, Surveying Users, and Documenting Facility Extent. FHWA Pedestrian and Bicycle Information Center, Chapel Hill, NC.

Sharma, S. C., Kilburn, P. & Wu, Y., 1996. The precision of average annual daily traffic estimates from seasonal traffic counts: Alberta example. *Can. J. Civ. Eng.*, (23), 302-304.

Yang, S. & Davis, G. A., 2002. Bayesian estimation of classified mean daily traffic. *Transportation Research Part A*, 365-382.

Zall Kusek, J. & Rist, R. C., 2004. Ten Steps to a Results-Based Monitoring and Evaluation System. The International Bank for Reconstruction and Development / The World Bank, Washington, DC.

Zegeer, C. V. et al., 2005. Safety Effects of Marked versus Unmarked Crosswalks at Uncontrolled Locations: Final Report and Recommended Guidelines. Office of Safety Research and Development, Federal Highway Administration, McLean, VA.

### 3. ROAD SAFETY PERFORMANCE MEASURES AND AADT UNCERTAINTY FROM SHORT-TERM COUNTS

The material in this chapter was submitted for publication in *Accident Analysis and Prevention* on August 18, 2014, and printed here with permission of co-authors Jeannette Montufar, and Jonathan Regehr; the journal editor has been notified. In following the paper, the chapter is self-contained with its own abstract, introduction, conclusion, and references. The author of this thesis had principal responsibility for all aspects of the paper, while co-authors provided advice and reviews.

**ABSTRACT: OBJECTIVE:** The objective of this paper is to enable better risk analysis of road safety performance measures by creating the first knowledge base on uncertainty surrounding annual average daily traffic (AADT) estimates when the estimates are derived by expanding short-term counts with the individual permanent counter method. **BACKGROUND:** Many road safety performance measures and performance models use AADT as an input. While there is an awareness that the input suffers from uncertainty, the uncertainty is not well known or accounted for. **METHOD:** The paper samples data from a set of 69 permanent automatic traffic recorders in Manitoba, Canada, to simulate almost 2 million short-term counts over a five year period. These short-term counts are expanded to AADT estimates by transferring temporal information from a directly linked nearby permanent count control station, and the resulting AADT values are compared to a known reference AADT to compute errors. The impacts of four factors on AADT error are considered: length of short-term count, number of short-term counts, distance from a count to its expansion control station, and the AADT at the count

site. **RESULTS:** The mean absolute transfer error for expanded AADT estimates is 6.7%, and this value varied by traffic pattern group from 5% to 10.5%. Reference percentiles of the error distribution show that almost all errors are between -20% and +30%. Error decreases substantially by using a 48-hour count instead of a 24-hour count, and only slightly by using two counts instead of one. Mean absolute transfer error increases with distance to control station (elasticity .121,  $p = .001$ ), and increases with AADT (elasticity .857,  $p < .001$ ). **IMPLICATIONS:** These results can support evidence-based risk analysis of road safety performance measures that use AADT as inputs. Analytical frameworks for such analysis exist but are infrequently used in road safety because the evidence base on AADT uncertainty is not well developed.

### 3.1. INTRODUCTION

A widely used input to road safety performance measures is annual average daily traffic (AADT), which analysts often estimate by applying temporal factors to expand short-term counts. At permanent count stations, AADT may also be calculated directly from the data. The uncertainty in AADT from short-term counts propagates into safety performance measures and is often overlooked.

This paper develops a new knowledge base on AADT uncertainty to enable better risk analysis of performance measures in road safety engineering. A sampling experiment uses permanent count station data to simulate approximately 2 million short-term counts along with corresponding expanded AADT estimates and errors to develop a new knowledge base on AADT uncertainty associated with using the

individual permanent counter (IPC) expansion method (Section 3.2.1 describes IPC). The paper also investigates four factors affecting the magnitude of uncertainty: the distance between a short-term count site and an expansion control station, the AADT (e.g. does AADT uncertainty change over a range of AADT values), number of short-term counts (e.g. error when expanding two counts instead of one), and length of short-term counts (e.g. error when expanding 48-hour counts instead of 24-hour counts).

Questions of risk and uncertainty in performance measures seem increasingly relevant to decision-makers in the current transport policy context (articulated, for example, in the U.S. MAP-21 legislation). To address these questions of risk and uncertainty, techniques such as Monte Carlo simulation modelling use information on input uncertainty to develop information on output uncertainty. However, the techniques depend on the availability of the evidence base on input uncertainty, and in the case of AADT, this evidence base is usually not available. As a result, the awareness that AADT uncertainty can cause road safety modelling and forecasting problems is more widespread than the practice of techniques to deal with these problems using a risk-based approach.

## **3.2. BACKGROUND AND NEED**

### **3.2.1. Estimating AADT from Short-Term Counts: Group Factors and Individual Permanent Counter Methods**

In basic traffic monitoring programs, analysts create AADT estimates at short-term count sites by expanding them with temporal data transferred from permanent

count sites (FHWA, 2013). Most United States and Canadian jurisdictions use the group factor method to make these adjustments (Yang & Davis, 2002) as does the United Kingdom (Government of the United Kingdom, 2005). Analysts using the group factor method arrange permanent and short-term counting stations into traffic pattern groups that exhibit similar temporal patterns. Methods for traffic pattern group creation and assignment include cluster analysis, land use attribute analysis, functional class grouping, engineering judgement (Li, Zhao, & Chow, 2006), Bayesian estimation (Yang & Davis, 2002), and hybrid methods (Reimer & Regehr, 2013). Once the traffic pattern groups are created, analysts use various averaging methods to calculate group day-of-week factors ( $D_h$ ) and month-of-year factors ( $M_h$ ) based on data from the permanent count sites in a group. This is represented in the FHWA Traffic Monitoring Guide (FHWA, 2013) formula for estimating AADT from a short-term count:

$$AADT_{hi} = VOL_i * M_h * D_h * A_i * G_h \quad (3-1),$$

where  $AADT$  is the expanded estimate,  $VOL$  is the short-term count volume expressed in axle counts,  $M$  and  $D$  are the month and day-of-week group expansion factors,  $A$  is an axle correction factor to convert axle counts to vehicle counts,  $G$  is a growth factor in the case of using old short-term counts to estimate  $AADT$ ,  $h$  represents a traffic pattern group, and  $i$  represents a site in that group.

Albright (1987) introduces the individual permanent count (IPC) method for expanding short term counts by transferring temporal information from an individual permanent count station that is considered as the ‘control station’ for that

short-term count station. Lucas (1996) describes the initial implementation of this method for Manitoba highways, where traffic pattern groups are first created by cluster analysis and then the individual permanent count control stations are selected from within these traffic pattern groups for each short-term count station. Poapst *et al.*, (2013, pp. 1-8) describe the current application of this method in Manitoba. Unlike the group factor method described above, which transfers general temporal information from a group, the IPC method transfers specific temporal information from one individual permanent count control site and from the same time period as the short-term count. In the IPC method, the equation to estimate AADT (adapted from Albright (1987) is:

$$AADT_{x,est} = V_{48,x} * AADT_{xc} / V_{48,xc} \quad (3-2),$$

where  $x$  is the short-term count site,  $xc$  is the volume control site (which is the closest permanent automatic traffic recorder (ATR) in the same traffic pattern group (TPG) as site  $x$ ), and  $V_{48}$  is a 48-hour volume during the short-term count period. In Manitoba, two 48-hour short-term counts are expanded in this way and the final AADT estimate is taken as the average of the two expanded estimates. In other applications of the method, the length or number of short-term counts may be different. Lucas (1996) cites the following benefits of the IPC method: (1) no need to re-calculate group factors every year; and (2) improved opportunity to capture localized spatial and temporal traffic patterns in the expansion process. The accuracy of this method is not well understood, and neither are the implications of this accuracy for safety performance measures that rely on AADT

as inputs, such as expected collisions, collision rates, countermeasure effectiveness estimates, or expected collision reductions from an intervention.

### **3.2.2. Past Studies on Estimating AADT Uncertainty from Expanded Short-Term Counts**

Table 3-1 summarizes previous studies that investigate accuracy of vehicle AADT estimates produced from expanded short-term counts. Recently, more focused studies have investigated uncertainties in expanded estimates of truck volumes (NCHRP, 2005) and expanded estimates of active transportation volumes (Nordback, Marshall, & Stolz, 2013) (Esawey, Lim, & Sayed, 2013) (Milligan, Poapst, & Montufar, 2013).



**Table 3-1: Past studies on vehicle AADT accuracy from expanded short-term counts**

<b>Study</b>	<b>Expansion method</b>	<b>Findings</b>	<b>Uncertainty measurement</b>
(Gadda, Kockelman, & Magoon, 2007)	Group factors, one 24-hr count	11.5% to 18.5%	Mean absolute percent error
(Jiang, McCord, & Goel, 2006)	Group factors, one 48-hr count	8%	Mean absolute relative error
(Hu, Wright, & Esteve, 1998)	Group factors, one 24-hr count	5.7% to 15.7%	Relative root mean square error
Sharma, Kilburn, and Wu (1995)	IPC; two, three, or four one-month counts per year	+/- 3.6%, +/-5.6%, and +/-8%	95 percent intervals for percent error for four, three, and two one-month counts respectively
(Bodle, 1966)	Group factors, various count lengths.	12.4%, 9.7%,9.5%,5.6%, and 5.7%	Relative root mean square error for one 24-hr, 24-hr-no-Friday, 48-hr, 5-day, and 7-day count respectively
(Granato, 1998)	Group factors, various count lengths, urban areas	6.7%, 6.3%, 6.3%, and 6.4%	Average absolute deviation for one 24-hr, 48-hr, 72-hr, and 96-hr count respectively
(Chen S. , 1981)	Group factors, one 48-hr count	7.3% to 13.9%	Standard deviation of relative errors

Some broad generalizations from Table 3-1 are that for expanded short-term vehicle counts, a typical error magnitude ranges from about 5 to 15 percent and that this error depends on the number and length of short-term counts expanded. Table 3-1 also shows that only one study has investigated IPC but it did so in the context of long seasonal counts and not the more typical short-term counts used in a conventional traffic monitoring program.

### **3.2.3. Treatment of AADT Uncertainty in Road Safety Performance**

#### **Measurement**

Hauer (2014) notes that uncertainty in AADT is one of the main difficulties in road safety regression modelling. While it is somewhat common for authors building road safety models to make some comment about the AADT uncertainty, it is less common for authors to explicitly account for this uncertainty, perhaps because the evidence base on this uncertainty or methods for dealing with it are not well understood. There are a few exceptions where the AADT uncertainty is explicitly considered:

- El-Bayouni and Sayed (2010) note that AADT uncertainty can bias estimates of effect and increase estimates of dispersion in SPFs. This is important in practical engineering terms because a main use of SPFs is to apply them to network screening in an Empirical Bayes approach where the weight assigned to the SPF prediction decreases with increasing dispersion. The authors implement what they call a ‘measurement error negative binomial’ (MENB) approach, which is a model including three equations: response, exposure, and measurement error. The MENB approach provided superior goodness of fit to the conventional negative binomial approach in the case of large measurement errors. The approach requires evidence on the variance to mean ratio (VMR) as a quantification of AADT uncertainty.

- Nordback, Marshall and Johnson (2014) developed bicycle SPFs for a U.S. city where the uncertainty in bicycle AADT estimates had been quantified. The authors used this uncertainty information to conduct a sensitivity analysis and found that bicycle SPF parameters were sensitive to the uncertainty in bicycle AADT.
- Maher and Summersgill (1996) outline and implement two approaches for dealing with errors in flow estimates. First, randomization experiments were used to vary the flow values according to a lognormal distribution with a variance to mean ratio of 10 percent. The approach showed that a bias of about 20 percent can be removed from parameter estimates when accounting for flow uncertainty in this way. Second, a formal functional model can be implemented to explicitly account for flow uncertainty in the log likelihood function; the approach showed that true standard errors may be appreciably larger than those estimated using a basic model that ignores flow uncertainty.

### **3.3. METHOD**

The research uses 5 years of continuous hourly traffic volume data from 69 permanent automatic traffic recorders (ATRs) on provincially operated highways in Manitoba, Canada. Regehr et al (2006) arrange the 69 ATRs into seven traffic pattern groups (TPGs) using a hybrid method combining cluster analysis and engineering judgement. Usually, a short-term count site is assigned to one of these TPGs and counts are expanded by transferring temporal information from the

closest ATR in the TPG. This research simulates short-term counts at an ATR where a reference AADT is known by sampling 24-hour and 48-hour periods from the continuous data. The research then expands the simulated short-term counts by transferring temporal information from the closest ATR in the same TPG. The expanded count can then be compared to the reference AADT to obtain error values. To generate a large set of sample errors for analysis, the research uses the following sequential simulation approach:

1. Select an ATR as site  $x$ ;
2. Identify the expansion control site  $x_c$  as the closest ATR in the same TPG;
3. Sample the continuous data at site  $x$  to simulate a set of short-term counts;
4. For each sampled short-term count, expand the count by using temporal data from  $x_c$  to estimate an AADT at site  $x$  (this process is applied individually and also applied and averaged across pairs of short-term counts to test the practice of using two short-term counts in a year);
5. Compute errors by comparing the estimate of AADT to the reference AADT;
6. Select another ATR as site  $x$  and repeat previous steps.

This research computes four types of errors for each expansion: transfer error ( $TE$ ), absolute transfer error ( $ATE$ ), relative transfer error ( $RTE$ ), and absolute relative transfer error ( $ARTE$ ). The errors are called transfer errors because they

are associated with the transfer of temporal information from a control site to the short-term count site. TE and RTE are defined as follows:

$$TE = AADT_{x,est} - AADT_x \quad (3-3);$$

$$RTE = ((AADT_{x,est}) - (AADT_x))/AADT_x \quad (3-4).$$

The method excludes cases where a transfer error cannot be calculated with the IPC method. While up to about 3 million error values are possible during the study period, there are cases where  $V_{48,x}$ ,  $V_{48,xc}$ ,  $AADT_x$ , or  $AADT_{xc}$  are not available because an ATR is out of service. In these cases, Manitoba uses the group factor method instead of the IPC method to estimate AADT. About 2 million error values are considered in the paper, all based on the IPC method.

The new knowledge base on AADT uncertainty using the IPC method is provided in the form of reference percentiles of the RTE distribution as well as the mean absolute relative transfer error, which can provide the required evidence to inform a risk-based approach to SPF development and application.

The investigation of factors contributing to uncertainty in expanded AADT estimates includes a comparative analysis of RTE distributions for select numbers of short-term counts (1 vs 2 per year) and lengths of short-term counts (24- vs 48-hour). The investigation also includes a regression analysis of mean absolute transfer error and mean absolute relative transfer error against two variables thought to impact the magnitude of these errors: the actual reference AADT and the distance to the control stations.

### 3.4. RESULTS

The results section describes summary characteristics of the database, the distributions of expansion errors, and regression outcomes showing the effect of volume and distance on mean error magnitude.

#### 3.4.1. Summary Characteristics of the Database

Table 3-2 shows summary characteristics of the database developed for this research. The main contribution of the database is the set of AADT transfer errors from expansions using the IPC method on a set of simulated short-term counts where a reference AADT value is known.

**Table 3-2: Database summary**

Parameter	Value
Number of Stations	69
Years of Data	5
AADT	Mean: 3932; SD: 4297
d	Mean: 47.3 km; SD: 50.1 km
TPGs	7
Number of AADT estimates and expansion errors	1,986,931

Notes: AADT = Annual Average Daily Traffic, d = distance to expansion control station; TPGs = traffic pattern groups

### 3.4.2. Distribution of Expansion Transfer Errors

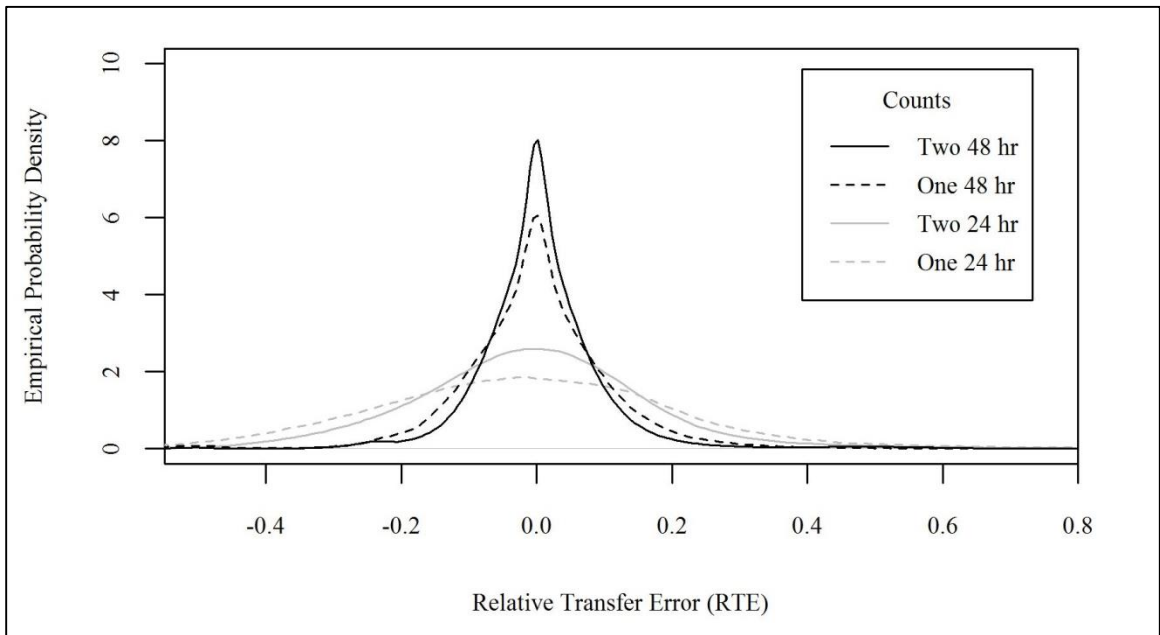
Table 3-3 shows seven-figure summaries for the distribution of relative transfer errors (RTE) as well as the mean absolute relative transfer error (MARTE). A seven-figure summary describes a distribution with seven reference percentiles – in this case, the 2nd, 9th, 25th, 50th, 75th, 91st, and 98th percentiles. The table gives overall results and results by traffic pattern group. The table shows an overall MARTE value of .067, with values across TPGs ranging from .050 to .105. This indicates a typical error magnitude of about +/- 6.7 percent. The 50th percentile RTE values near zero indicate little to no bias in the expansion method. Finally the 9th and 91st percentiles indicate that over 80 percent of errors are limited to the range of about +/- 10 percent.

**Table 3-3: Distribution of transfer errors when using individual permanent counter method to expand two 48-hour short-term counts**

TPG	n	MARTE	RTE.02	RTE.09	RTE.25	RTE.50	RTE.75	RTE.91	RTE.98
PG1	216732	0.068	-0.187	-0.102	-0.043	0.001	0.048	0.123	0.272
PG2	787759	0.073	-0.217	-0.094	-0.037	0.002	0.044	0.114	0.467
PG3	120403	0.059	-0.196	-0.088	-0.036	0.008	0.048	0.094	0.160
PG4	442817	0.058	-0.156	-0.088	-0.041	0.000	0.043	0.100	0.196
PG5	96431	0.105	-0.305	-0.126	-0.058	-0.007	0.043	0.121	0.634
PG6	171437	0.065	-0.176	-0.111	-0.056	-0.003	0.050	0.104	0.171
PG7	151352	0.050	-0.161	-0.083	-0.020	0.001	0.024	0.110	0.216
ALL	1986931	0.067	-0.194	-0.095	-0.040	0.001	0.044	0.108	0.289

Notes: TPG = traffic pattern group; MARTE = mean absolute relative transfer error; RTE.n = n<sup>th</sup> percentile of the relative transfer error distribution. TPG definitions given in background section. Errors are based on comparing an AADT estimate obtained by expanding two 48-hour short-term counts at a site to the reference AADT value for that site.

Figure 3-1 shows empirical probability density functions for the relative transfer error based on the length and number of short-term counts expanded to estimate AADT. The “Two 48-hr” curve corresponds to Table 3-3 above and to the practice in Manitoba of estimating AADT at a short-term count site by conducting two 48-hour counts at different points in the year, expanding them both using the IPC method, and taking the average of both expansions. Additional curves show the impact on the relative transfer error distribution of expanding only one count or limiting the short-term count length to 24-hours. The results show a substantial improvement for expanding 48-hour counts instead of 24-hour counts, and a moderate improvement for expanding two counts instead of one.



**Figure 3-1: RTE distribution by number and duration of expanded short-term counts**



### **3.4.3. Regressions of Transfer Errors on AADT and Distance to Expansion Control Station**

An understanding of the factors influencing the magnitude of transfer errors can assist practitioners in counting program design and can also provide guidance in adjusting risk analysis for specific scenarios. For example, if it is known that RTE is lower for higher AADT values, a practitioner may use lower RTE values than those given in Table 3-3 for a risk analysis if the application context is a facility with high AADT.

The regressions consider two explanatory variables thought to influence error magnitude: facility AADT and the distance between the short-term count site and the expansion control station (exploratory regressions also considered the effect of year on error, finding the year effects to be not meaningful and/or not significant). The regressions consider the impact of these variables on two dependent variables: mean absolute transfer error (MATE), and mean absolute relative transfer error (MARTE). In both cases, the mean refers to the mean by site and year. For example, for a given count station and year, the database may contain 6000 simulated errors, and these 6000 errors would yield one MATE and one MARTE for the regression dataset. The almost 2 million error values in the knowledge-base yield 240 means by site and year as records in the regression dataset. Figure 3-2 shows natural and log-transformed scatterplots of the dependent and independent variables. These scatterplots are used to help select the regression model specification. In the transformed scatterplots, error magnitude appears to increase strongly with AADT, while relative error appears to

decrease slightly with AADT. Error magnitude and relative errors appear to increase slightly with the distance to the expansion control station. Because these relationships are not as evident or do not appear linear in the untransformed dataset, the log-transformed relationships are used in the following regression model specification, where  $\beta$  and  $\varepsilon$  represent coefficients and normal random errors, respectively:

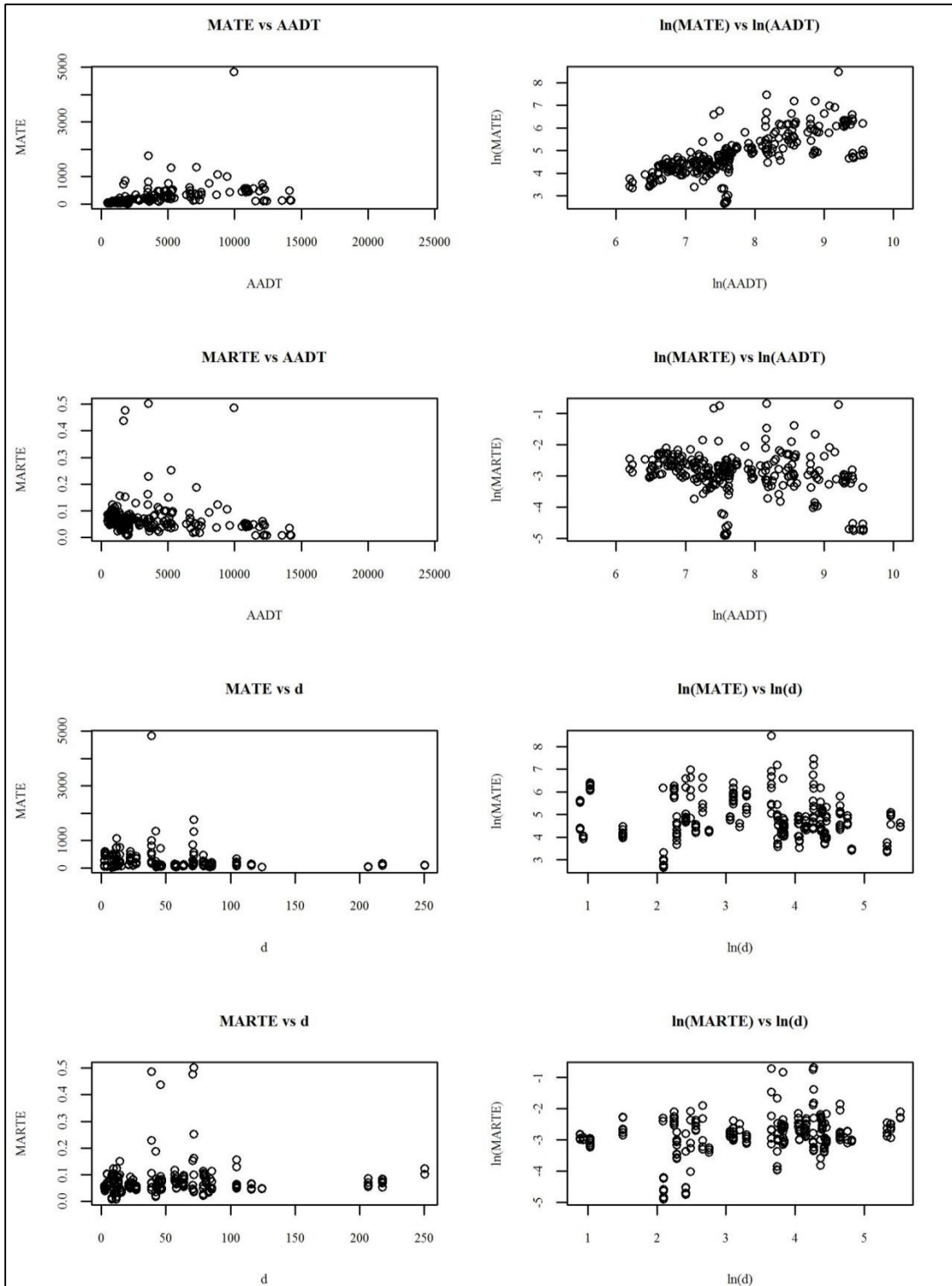
$$\ln(MATE) = \beta_0 + \beta_1 * \ln(AADT) + \beta_2 * \ln(d) + \varepsilon \quad (3-5)$$

$$\ln(MARTE) = \beta_0 + \beta_1 * \ln(AADT) + \beta_2 * \ln(d) + \varepsilon. \quad (3-6)$$

The two ordinary least squares linear regression models were estimated in R software (R Core Team, 2013).

Table 3-4 gives the regression results and reveals that AADT and the distance from the short-term count site to the control station have statistically significant relationships with error values. The model coefficients have a natural interpretation as the elasticity of the error value with respect to the explanatory variable. MATE errors (not normalized to AADT) increase with AADT at an elasticity of .86 and increase with distance at an elasticity of .12. MARTE errors (normalized to AADT) decrease slightly with AADT at an elasticity of -.14, and increase slightly with distance at an elasticity of .12; all of these relationships are statistically significant with p-values below .01. The MATE model has reasonable explanatory power with  $r^2_{adj}$  of .552 while the MARTE model has low explanatory power with  $r^2_{adj}$  of .111

. This indicates that once the errors are normalized to volume, most of the remaining variation in errors is either random or due to unobserved factors, with a small portion still explained by volume (relative error decreases) and by distance (relative error increases). An assumption in ordinary least squares modeling is that the model residuals (difference between observed and fitted values) should be relatively randomly distributed across the ranges of explanatory variables. As a check on this assumption, **Figure 3-3** gives the model residuals against explanatory variables, revealing no unusual patterns.

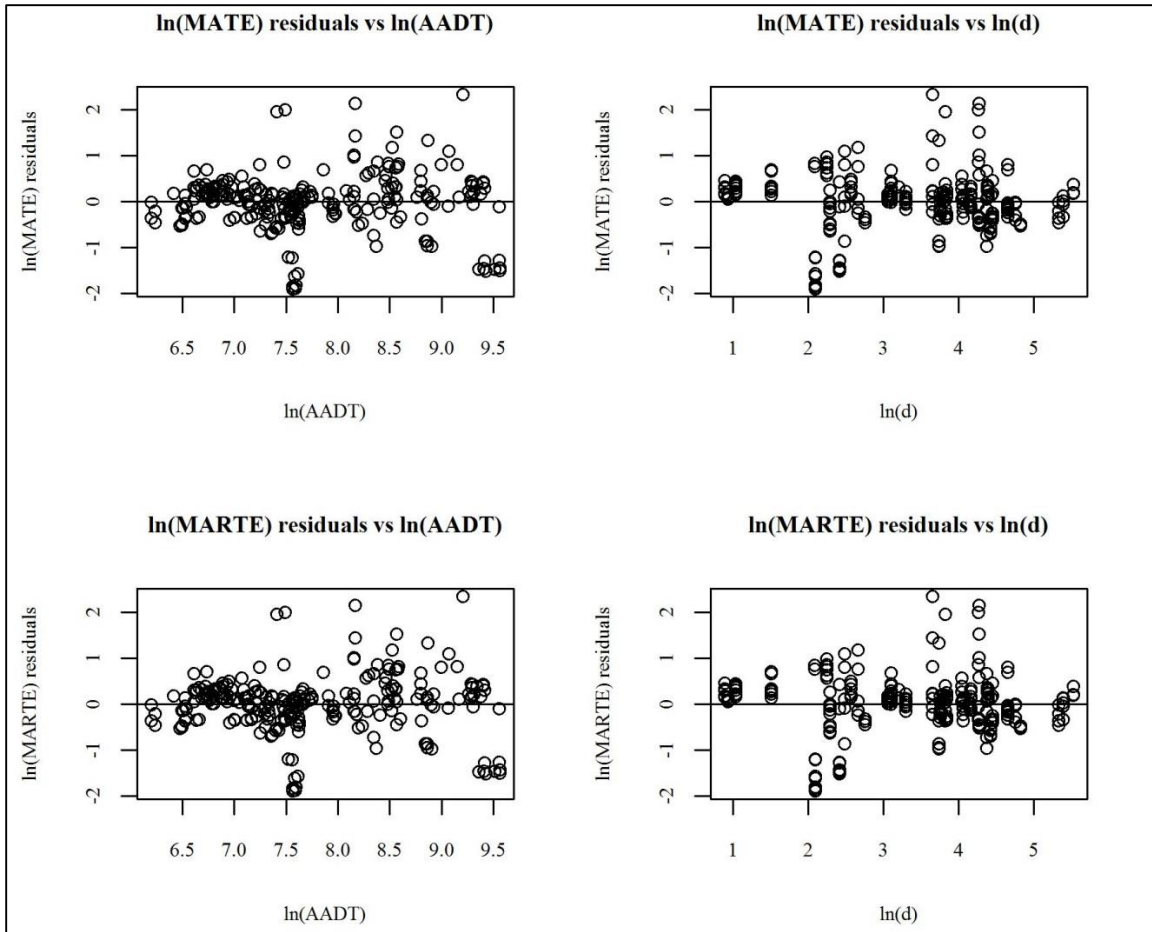


**Figure 3-2: Scatterplots of errors against potential explanatory variables.**  
 Notes: MATE = Mean Absolute Transfer Error (by count site and year), MARTE = Mean absolute relative transfer error (by count site and year), AADT = Annual Average Daily Traffic, d = distance from short-term count site to expansion control station (in km).

**Table 3-4: Regression of error values on volume and distance to control station**

	ln(MATE)			ln(MARTE )		
	(n=240, $r^2_{adj} = .552$ )			(n=240, $r^2_{adj} = .111$ )		
	$\beta$	SE	p	$\beta$	SE	p
Intercept	-2.181	0.455	3E-06	-2.181	0.455	3E-06
ln(AADT)	0.857	0.051	8E-43	-0.143	0.051	0.005
ln(d)	0.121	0.037	0.001	0.121	0.037	0.001

Notes: MATE - model predicts ln(mean absolute transfer error); MARTE - model predicts ln(mean absolute relative transfer error);  $\beta$  = coefficient estimate; SE = standard error of coefficient estimate; AADT = Annual Average Daily Traffic; d = distance from short-term count site to expansion control station; n = the number of mean errors in the regression where the mean is by site and by year.



**Figure 3-3: Residual plots for regressions of error values against explanatory variables.** Notes: MATE = mean absolute transfer error (by site and year); MARTE = mean absolute relative transfer error (by site and year); AADT = annual average daily traffic; d = distance from short-term count site to expansion control station, in km.

### 3.5. DISCUSSION AND CONCLUSION

While AADT is acknowledged as uncertain and the parameter is heavily leaned on as an input to safety performance measures, explicit treatment of uncertainty of AADT in safety performance estimates is not common. This study represents, to the authors' knowledge, the first empirical information on AADT uncertainty when AADT is estimated by expanding short-term counts using the IPC method. This in turn can allow explicit treatment of AADT uncertainty in safety performance

measures that depend on AADT as an input. The distribution of relative transfer errors yielded by the research shows that the IPC expansion method is unbiased, with median relative transfer errors of less than one percent. The overall mean absolute relative transfer error of 6.7% can be a useful figure for a discrete sensitivity analysis of the influence of AADT uncertainty on safety performance measure uncertainty. More extreme values of the distribution – for example the 2<sup>nd</sup> and 98<sup>th</sup> percentile RTE values of -19% and +29% can be used for a more comprehensive probabilistic risk analysis that uses Monte Carlo simulation (for example, as in the type of risk analysis used in the *Road Economic Decisions* software of the World Bank (Archondo-Callao, 2004) or other similar risk based software). These risk analysis values are relevant to the context investigated in this paper (short-term counts expanded in Manitoba using the IPC method), while the method provided in the paper can be used to obtain analogous risk analysis values for other contexts.

In addition to characterizing AADT uncertainty for risk analysis, the paper also investigated factors that may impact the magnitude of this uncertainty. The investigation revealed two statistically significant relationships. First, while AADT errors increase with AADT, they do so at an elasticity of .86, so that the relative or percentage error decreases slightly with AADT. Second, AADT errors increase slightly (elasticity .12) with the distance from the short-term count site to the expansion control station, meaning that when temporal information is transferred from a control station further away, errors are slightly higher. A comparative analysis of RTE distributions with varying count length and duration showed lower

RTE for 48-hour counts as opposed to 24-hour counts, and lower RTE for taking the average of two expanded counts as opposed to using only one expanded count. However, the uncertainty reduction from using two counts is not large and should be weighed against the cost of counting locations twice.

This paper has some natural limitations related to its scope. In other contexts, where traffic patterns are different, or methods to create traffic pattern groups or to expand short-term counts are different, the resulting analysis of AADT uncertainty will be different. However, most jurisdictions that have a short-term count program linked to a permanent count program for expansion purposes could apply the methodology in this paper to assess the resulting uncertainty in AADT estimates for safety performance purposes.

The practice of performance-based decision-making in road safety engineering is continually progressing with the quality of safety performance information. The dissemination and adoption of safety performance analysis tools such as those in the Highway Safety Manual (AASHTO, 2010) or in the Collision Modification Factors Clearinghouse (FHWA, 2013) has enabled improved performance-based decision making in road safety. However, the usefulness of these tools can be limited and results questioned when the tools rely on inputs for which the uncertainty is unknown. With an expanded knowledge base on AADT uncertainty, safety performance measurement tools can be applied within a risk-based approach.



### **3.6. ACKNOWLEDGEMENTS**

The authors gratefully acknowledge use of Manitoba Highway Traffic Information System datasets, literature search efforts by Mr. Adam Budowski, EIT, and financial support from the Vanier Canada Graduate Scholarship Program and the Edward Toporek Graduate Fellowship in Engineering (University of Manitoba).

### **3.7. REFERENCES**

AASHTO. (2010). *Highway Safety Manual* (First ed.). Washington, DC: AASHTO.

Albright, D. (1987). A Quick Cluster Control Method: Permanent Control Station Cluster Analysis in Average Daily Traffic Calculations. *Transportation Research Record*, 1134, 57-64.

Archondo-Callao, R. (2004). *Roads Economic Decision Model Software User Guide and Case Studies* (SSATP Working Paper No. 78). Washington, DC: The World Bank, Africa Region.

Bodle, R. (1966). *Evaluation of Rural Coverage Count Duration for Estimating Annual Average Daily Traffic*. Washington, DC: U.S. Department of Commerce Bureau of Public Roads.

Chen, S. (1981). *Improved Methods for State-Wide Vehicle Counting*. West Lafayette, Indiana: Purdue University Joint Highway Research Project.

El-Bayouny, K., & Sayed, T. (2010). Safety performance functions with measurement errors in traffic volume. *Safety Science*, 48, 1339-1344.

Esawey, M., Lim, C., & Sayed, T. M. (2013). Development of daily adjustment factors for bicycle traffic. *Journal of Transportation Engineering*, 139(8), 859-871.

FHWA. (2013). *Crash Modification Factors Clearinghouse*. Retrieved 01 11, 2013, from <http://www.cmfclearinghouse.org/>

FHWA. (2013). *Traffic Monitoring Guide*. Washington, DC: United States Federal Highways Administration.

Gadda, S., Kockelman, K., & Magoon, A. (2007). Estimates of AADT: Quantifying the Uncertainty. *World Conference on Transportation Research* (pp. 1-17). Berkley, CA: WCTR.

Government of the United Kingdom. (2005). How the National Road Traffic Estimates are made. *Workshop on statistics on the volume of road traffic (vehicle-kilometres)* (pp. 1-6). Copenhagen: United Nations Economic Commission for Europe.

Granato, S. (1998). The Impact of Factoring Traffic Counts for Daily and Monthly Variation in Reducing Sample Counting Error. *Proceedings of the Crossroads 2000 Conference* (pp. 122-125). Ames, Iowa: Iowa State University Center for Transportation Research and Education.

Hauer, E. (2014). *The art of regression modeling in road safety* (January 2014 ed.). Toronto, Canada: Ezra Hauer.

Hu, P. S., Wright, T., & Esteve, T. (1998). Traffic Count Estimates for Short-Term Traffic Monitoring: Simulation Study. *Transportation Research Record*, 26-34.

Jiang, Z., McCord, M. R., & Goel, P. K. (2006). Improved AADT Estimation by Combining Information in Image- and Ground-Based Traffic Data. *Journal of Transportation Engineering*, 523-530.

Li, M.-T., Zhao, F., & Chow, L.-F. (2006). Assignment of Seasonal Factor Categories to Urban Coverage Count Stations Using a Fuzzy Decision Tree. *Journal of Transportation Engineering*, 654-662.

Lucas, B. (1996). *Design, Development, and Implementation of the Manitoba Highway Traffic Information System (Master's Thesis)*. Winnipeg: University of Manitoba.

Maher, M. J., & Summersgill, I. (1996). A comprehensive methodology for the fitting of predictive accident models. *Accident Analysis and Prevention*, 281-296.

Milligan, C., Poapst, R., & Montufar, J. (2013). Performance measures and input uncertainty for pedestrian crossing exposure estimates. *Accident Analysis and Prevention*, 50, 490-498.

NCHRP. (2005). *NCHRP Report 538: Traffic Data Collection, Analysis, and Forecasting for Mechanistic Pavement Design*. Transportation Research Board.

Nordback, K., Marshall, W. J., & Stolz, E. (2013). Estimating annual average daily bicyclists. *Transportation Research Record*, 2339, 90-97.

Nordback, K., Marshall, W., & Janson, B. (2014). Bicyclist safety performance functions for a U.S. city. *Accident Analysis and Prevention*, 114-122.

Poapst, R., Budowski, A., Vogt, M., & Montufar, J. (2013). *Traffic on Manitoba Highways 2012*. Winnipeg, MB: Manitoba Highway Traffic Information System.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Australia: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>

Regehr, J., Rådström, B., Arango, J., Isaacs, C., Han, K., Rempel, G., . . . Clayton, A. (2006). *Manitoba Highway Traffic Information System: System Maintenance and Upgrade*. Winnipeg, MB: University of Manitoba Transport Information Group.

Reimer, M., & Regehr, J. (2013). Hybrid approach for clustering vehicle classification data to support regional implementation of the mechanistic-empirical pavement design guide. *Transportation Research Record*, 2339, 112-119.

Yang, S., & Davis, G. A. (2002). Bayesian estimation of classified mean daily traffic. *Transportation Research Part A*, 365-382.

#### 4. VALUE OF A STATISTICAL LIFE IN ROAD SAFETY: A BENEFIT-TRANSFER FUNCTION WITH RISK-ANALYSIS GUIDANCE BASED ON DEVELOPING COUNTRY DATA

The material in this chapter is published in (Milligan et al., 2014), and reprinted with permission of co-authors Andreas Kopp, Said Dahdah, and Jeannette Montufar. In following the paper, the chapter is self-contained with its own abstract, introduction, conclusion, and references. The author of this thesis had principal responsibility for all aspects of the paper, while co-authors provided advice and reviews.

**ABSTRACT:** We model a value of statistical life (VSL) transfer function for application to road-safety engineering in developing countries through an income-disaggregated meta-analysis of scope-sensitive stated preference VSL data. The income-disaggregated meta-analysis treats developing country and high-income country data separately. Previous transfer functions are based on aggregated datasets that are composed largely of data from high-income countries. Recent evidence, particularly with respect to the income elasticity of VSL, suggests that the aggregate approach is deficient because it does not account for a possible change in inter-country income elasticity across income levels. Our dataset (a minor update of the OECD database published in 2012) includes 123 scope-sensitive VSL estimates from developing countries and 185 scope-sensitive estimates from high-income countries. The transfer function for developing countries gives  $VSL = 1.3732E-4 \cdot (GDP \text{ per capita})^{2.478}$ , with  $VSL$  and  $GDP \text{ per capita}$  expressed in 2005 international dollars (an international dollar being a

notional currency with the same purchasing power as the U.S. dollar). The function can be applied for low- and middle-income countries with GDPs per capita above \$1268 (with a data gap for very low-income countries), whereas it is not useful above a GDP per capita of about \$20,000. The corresponding function built using high-income country data is  $VSL = 8.2474E+3*(GDP \text{ per capita})^{.6932}$ ; it is valid for high-income countries but over-estimates VSL for low- and middle-income countries. The research finds two principal significant differences between the transfer functions modeled using developing-country and high-income-country data, supporting the disaggregated approach. The first of these differences relates to between-country VSL income elasticity, which is 2.478 for the developing country function and .693 for the high-income function; the difference is significant at  $p < 0.001$ . This difference was recently postulated but not analyzed by other researchers. The second difference is that the traffic-risk context affects VSL negatively in developing countries and positively in high-income countries. The research quantifies uncertainty in the transfer function using parameters of the non-absolute distribution of relative transfer errors. The low- and middle-income function is unbiased, with a median relative transfer error of -.05 (95% CI: -.15 to .03), a 25<sup>th</sup> percentile error of -.22 (95% CI: -.29 to -.19), and a 75<sup>th</sup> percentile error of .20 (95% CI: .14 to .30). The quantified uncertainty characteristics support evidence-based approaches to sensitivity analysis and probabilistic risk analysis of economic performance measures for road-safety investments.

#### 4.1. INTRODUCTION

Analyses of investments to prevent road fatalities often use the net present value (NPV), the internal rate of return (IRR) or the social benefit-cost ratio as a prospective transport performance measure. These performance measures require estimates of both the value of a statistical life (VSL) and the value of a statistical injury (VSI). A robust and conservative engineering economic analysis using these performance measures also requires estimates of uncertainty in VSL and VSI. Many developing countries do not have appropriate VSL estimates and need to adapt existing estimates from elsewhere using transfer functions in a process called benefit-transfer. The currently available benefit-transfer functions are based on meta-analyses of datasets composed primarily of high-income country data, which may not be appropriate for application in developing countries.

The objectives of this research are to (1) develop a new VSL transfer function for application to transport safety in developing countries that is based on VSL estimates from developing countries, (2) determine whether this function differs significantly from functions that are based on VSL estimates from developed countries and (3) quantify the uncertainty associated with this new transfer function for practical application to the risk analysis of performance measures.

The study accomplishes these objectives by performing a new meta-analysis on a database of VSL estimates that has been made available as an accompaniment to the publication *Mortality Risk Valuation in Environment, Health and Transport Policies* (OECD 2012). Meta-analysis, which is widely used in road safety and

other fields of research, is “a quantified synthesis of the results of several studies” (Elvik, Høye, Vaa, & Sørensen, 2009, p. 20). The research also expands on the existing techniques for transfer error analysis and interpretation to validate the transfer function and enable its application in a stochastic framework.

The work is a subset of a project at the World Bank to develop a flagship report entitled *Comprehensive Assessment of Transport Policies and Projects* that will provide *ex ante* evaluation instruments to allow engineers to incorporate wider, multi-sectoral benefits of transport as well as environmental and safety costs into decision-making supports.

#### **4.2. EXISTING KNOWLEDGE, PRACTICES, AND NEEDS**

This section is organized into four subsections. Section 4.2.1 presents the general need for VSL estimates as inputs to the social benefit-cost analysis of road safety investments. Section 4.2.2 provides an overview of the methods used to create original VSL estimates along with their strengths and weaknesses. Section 4.2.3 describes the process of transferring VSL estimates to policy contexts in which no appropriate original VSL estimate exists and the current practice for assessing the uncertainty related to these transfers. Finally, Section 4.2.4 describes the state of existing practice for obtaining VSL estimates in developing countries and the emergence of opportunities to improve the state of this practice.



#### **4.2.1. The Transport Safety Problem and the Need for VSL Estimates in Benefit-Cost Analysis**

The need for this research is fundamentally predicated on the transport safety problem in developing countries, which has the dimensions of a global disease. While transport risks to individual users may appear low, the cumulative impact of these risks places a high burden on society. Nordfjærn *et al.* (2012) describe the problem as “increasing towards endemic proportions in developing countries” (p.1862). Worldwide, there are approximately 1.3 million road transport fatalities per year—or approximately 3500 per day (WHO, 2012). Analysts expect these rates to increase, and developing countries bear a high share of the burden (World Bank and WHO, 2004). Because of the magnitude of the problem and in recognition of health-related millennium development goals, the World Bank focuses on safety as the first of three themes in its transport business strategy for 2008 to 2012, entitled *Safe, Clean, and Affordable Transport for Development* (World Bank, 2008).

Many engineering countermeasures—in the form of policies or projects—are available to reduce the risk of transport fatalities and injuries. Elvik *et al.* (2009) review the expected effectiveness levels of various countermeasures, as do several other handbooks and toolkits. With the resulting estimated changes to physical indicators in hand (i.e., reductions in fatalities or serious injuries), governments turn to social benefit-cost analysis (BCA) to develop performance measures that evaluate transport safety spending vis-à-vis other potential public spending from the perspective of overall welfare. An in-depth guide to project

evaluation using social BCA is provided by Dasgupta, Sen and Marglin (1972). Market prices often provide suitable information about public preferences for use in BCA, but in many cases, they do not. In these cases, social BCA requires the use of shadow prices, which are notional prices for the physical costs and benefits used by the government to reflect public preferences for evaluation purposes (Dasgupta, Sen, & Marglin, 1972). When social BCA addresses transport safety, shadow prices are required for the benefits of reduced transport risks because no market directly deals in these benefits. Most work to develop shadow prices for road safety produces a VSL or a VSI. The costs of property damage only (PDO) collisions are more amenable to evaluation at market prices because there are functioning markets that deal in the repair or replacement of damaged property (namely, vehicles). Furthermore, although the PDO costs are significant, they are small compared with the costs of injuries and fatalities. It is important to note that the VSL values do not reflect the moral value of a person's life. An appropriate VSL value is one that supports social BCA by reflecting the preferences of individual members of the public related to their individual marginal rates of substitution between risk and income. Although social BCA is a widely used tool to evaluate road safety investments according to public preferences, it is not the only approach. Other approaches to evaluate road safety investments include cost-effectiveness analysis, vision zero (see support in (Rosencrantz, Edvarsson, & Hansson, 2007) and criticism in (Elvik, 1999)), multi-criteria analysis (e.g., an impact tableau (Manheim, 1979)), and citizen's juries (see arguments in favor by (Hauer E. , 2011)). Although some researchers prefer and argue for these other

approaches, this paper develops a new VSL benefit transfer function for application to road safety BCA in developing countries—though alternative approaches to BCA exist—under the assumption that the conventional practice of social BCA will continue for some time and that social BCA is useful for evaluation purposes.

#### **4.2.2. Methods to Estimate VSL and their Strengths and Weaknesses**

The methods used to estimate the value of a statistical life fall into two categories: the human capital (HC) method and the willingness to pay (WTP) method. The HC method uses lost productivity calculations, and analysts have almost completely abandoned this method because it fails to account for intangible dimensions, such as suffering and grief. They instead favor the WTP method, which implicitly includes these dimensions and is based on consumer preferences, which form the basis of BCA under the new welfare economics paradigm. The WTP method is further classified into two categories: the stated preference (SP) and revealed preference (RP) methods.

The stated preference (SP) method uses surveys that are designed to elicit from participants a statement about the quantity of money that they would be willing to spend to achieve a small reduction in mortality risk. These surveys are based on an assumption that individuals can state their real preferences regarding a marginal rate of substitution between wealth and a specific type of mortality risk reduction when asked hypothetical questions about these preferences. If a person states a willingness to pay \$10,000 towards a policy that will reduce their risk of

dying from 1.5% to .5%, the value of a statistical life is calculated as the willingness to pay divided by the risk reduction, or \$10,000 divided by 1%, giving VSL = \$1 million. If there were 100 identical people, the expected number of deaths reduced by implementing the policy for the entire group is 1 (reduced from 1.5 to .5), and as a group, the total willingness to pay to save that one statistical life is \$1 million (100 people times \$10,000).

The revealed preference (RP) method observes behavior in a proxy market to measure the actual willingness to pay for small reductions in mortality risk; the method then calculates the VSL in the same way that the SP method uses. The RP method requires a market for behavior observation. The RP method uses two markets: (1) the labor market and (2) the market for risk-reducing consumer goods (OECD, 2012) (Hauer E. , 2011) (Kochi, Hubbell, & Kramer, 2006) (Viscusi & Aldy, The value of a statistical life: a critical review of market estimates throughout the world, 2003) (Miller, 2000). The labor market method attempts to identify the wage premium associated with risk in jobs. The risk-reducing consumer goods market method attempts to identify the WTP for risk reduction by combining prices for protective goods—such as smoke alarms or safer cars—with estimates of the corresponding risk reductions offered by these goods. This WTP reflects a lower-bound WTP for the buyers of the good and an upper-bound WTP for the non-buyers of the good. Within RP methods, the wage-risk market is used more frequently than the risk-reducing consumer goods market.

Among developed countries, the United States has traditionally relied on RP methods using the labor market for VSL estimates, whereas European countries,

Canada, and Australia tend to use the SP approach (OECD, 2012). Overall, there is a growing emphasis on SP methods (OECD, 2012), and the use of RP methods has slowed significantly since 1990 (Miller, 2000).

Table 4-1 summarizes the literature on the strengths and weaknesses of the SP and RP methods for obtaining VSL estimates for use in engineering economics. The main strengths of the SP method are its ability to match survey questions to the policy risk context and achieve broad representation through survey design and control. The main strength of the RP method is its basis on actual behavior.

Table 4-1 indicates that one of the weaknesses of RP wage-risk market studies is the lack of equilibrium in the employment market because of the high transaction costs associated with changing jobs, which results in an upward bias in RP-based VSL estimates. For a given risk difference between two jobs, the wage premium required to switch jobs is thus higher than the actual risk premium because of the transaction costs. Because this method bases VSL on the ratio of the wage premium to the risk difference, the transaction cost effectively upwardly biases the VSL estimate. Another factor that can introduce an upward bias in the RP wage-risk methods is a tendency of regression modelers to remove workplace injury risk from the regression models to avoid any associated multicollinearity issues, with the result that the calculated wage premium for mortality risk is actually the premium for mortality and injury risk together (Miller, 2000). This theoretical upward bias is also demonstrated empirically: Kochi *et al.* (2006) compare distributions of VSL estimates based on SP and RP methods and find that RP estimates are higher and more dispersed than SP estimates.

**Table 4-1: Strengths and weaknesses of the stated and revealed preference methods**

Stated Preference		Revealed Preference	
Strengths	Weaknesses	Strengths	Weaknesses
-Flexibility to control for many variables including risk context <sup>a</sup>	-Based on hypothetical behaviour <sup>a,b</sup>	-Based on actual behaviour <sup>a,c</sup>	-Context-insensitive, but risk valuation is context-sensitive <sup>a,d,e,f</sup>
-Can elicit preferences for non-observable attributes <sup>g</sup>	-Lack of systematic responses to very small risk changes <sup>a,f,g,h</sup>	-Some research finds consensus that wage is responsive to risk <sup>i</sup>	-Some research finds that the wage-risk relationship is spurious <sup>e,j</sup>
-Can be representative of population if well designed <sup>f</sup>			-Difficult to account for non-risk determinants of wage variation <sup>a,b,f,k</sup>
			-Panel data only gives cross-individual rates of substitution <sup>d</sup>
			-High transaction costs means that workers not at wage-risk equilibrium <sup>h</sup>
			-Wage-earners are not representative of the population <sup>f</sup>
			-Estimates are distorted by the gap between real and perceived risks <sup>a,f,h</sup>

Notes: <sup>a</sup> (Hauer E. , 2011); <sup>b</sup> (Kochi, Hubbell, & Kramer, 2006); <sup>c</sup> (Cnaan & Kang, 2011), <sup>d</sup> (Grüne-Yanoff, 2009); <sup>e</sup> (Mrozek & Taylor, 2002) <sup>f</sup> (OECD, 2012) <sup>g</sup> (Goldberg & Roosen, 2007); <sup>h</sup> (McConnel, 2006); <sup>i</sup> (Ruser & Butler, 2009); <sup>j</sup> (Miller, 2000); <sup>k</sup> (Viscusi & Aldy, 2003).

The main weakness of the SP method, stemming from its hypothetical basis, is that SP surveys sometimes fail to elicit systematic responses to very small risk changes. People have difficulty comprehending small numbers and small risk changes (Hauer E. , 2011), and they may sometimes respond with the same WTP for risk changes that differ significantly.

Although both methods have weaknesses, there are opportunities to address these weaknesses by applying the methods within a careful design. For example, in SP surveys, the use of visual aids to help participants understand risk changes greatly reduces the variability in the resulting VSL estimates (OECD, 2012). In RP wage-risk studies, the use of multivariate regression modelling can be employed to isolate risk factors from other determinants of wages.

This paper uses VSL estimates developed using SP methods to model transfer functions, principally due to the method's ability to investigate risk context. There is some degree of uncertainty in all VSL estimates, and this should be quantified and accounted for to support robust economic assessments to the greatest extent possible.

#### **4.2.3. Transferring VSL Estimates Between Policy Contexts: Needs, Methods, and Uncertainty**

Many countries do not have SP-based VSL (Dahdah & McMahon, 2008) (Miller, 2000). When a social BCA is required to support engineering economic analysis in a policy context without an appropriate VSL estimate, analysts use benefit-transfer to obtain an estimate for the policy context based on estimates that have been derived elsewhere. The OECD (2012) gives details on five methods to conduct benefit-transfer. In order of the amount of information incorporated into the transfer process, these methods are: (1) simple (naïve) unit value transfer; (2) unit value transfer with income adjustments; (3) unit value transfer for separate age groups; (4) benefit function transfer from a single study; and (5) benefit

function transfer by meta-analysis. This paper uses benefit function transfer by meta-analysis, which incorporates the greatest amount of information in the transfer process.

The quality of a benefit transfer function and the uncertainty associated with using it are measured using a process called transfer error analysis. The OECD (2012) describes this process as follows, drawing on Navrud and Ready (2007), Kristofersson and Navrud (2005), and Kristofersson and Navrud (2007). For one application of a transfer function where the result of the transfer function,  $VSL_{TF}$ , can be compared with a benchmark value,  $VSL_B$ , the transfer error is commonly defined as

$$TE = 100\% * |VSL_{TF} - VSL_B| / VSL_B. \quad (4-1)$$

After many comparisons are made and a set of  $TE$  values is generated, previous studies have summarized transfer error performance using the mean or median  $TE$  values. The OECD (2012) tests the transfer function results using the actual estimates in the database as benchmark values. This allows as many tests of the transfer function as there are estimates in the database.

Previous studies (e.g., OECD 2012) summarize benefit transfer performance using the mean or median absolute transfer error. A summary of benefit transfer performance using the mean of absolute transfer errors has some limitations—namely, that this measure (1) is actually a description of the dispersion of non-absolute errors, (2) does not account for bias or asymmetry in the data, and (3) is highly sensitive to outliers. Some past studies have suggested ways to expand the



approach to transfer error analysis. For example, Lindhjem and Navrud (2008) note the influence of TE outliers in the discussion of their Figure 3, prompting a focus on the 40<sup>th</sup>, 50<sup>th</sup>, and 70<sup>th</sup> percentiles of the absolute TE distribution. Rosenberger and Loomis (2000) show both absolute and non-absolute values of percent errors in their Table 5; the ranges of values that they obtain for the absolute errors show a clear asymmetry around zero of the error data, which in turn suggests that the use of absolute errors leads to some information loss.

#### **4.2.4. VSL and VSI in Developing Countries**

All of the transfer functions used to estimate VSL for a developing country rely on evidence or assumptions about the between-country income elasticity of VSL. This elasticity is a ratio of the percentage difference in VSL between two countries to the percentage difference in incomes between two countries. Between-country income elasticity may differ from within-country income elasticity; consequently, generalizations from one to the other can be misleading. This paper focuses on between-country income elasticity rather than within-country income elasticity because the former is the elasticity that is relevant to transfer functions.

In early work on VSL transfer to developing countries, Miller (2000) develops a set of preliminary VSL transfer functions, but they are based on a regression of estimates from 13 primarily high income countries. Miller (2000) proposes that the functions could provide reasonable estimates for developing countries but notes an urgent need for more research to investigate the ability of these functions to predict VSL for lower-income countries that are beyond the range of the source

data for the functions. The functions, built using a database composed largely of high-income country VSL estimates, indicate an income elasticity for VSL between .85 and 1.0.

Viscusi and Aldy (2003) provide a meta-analysis based on 60 estimates from wage-risk studies. The study uses a database composed largely of high-income country estimates, although four of the studies in the meta-analysis are based on developing countries. They find an income elasticity of VSL in the range of .5 to .6. They note that the regression results for the income elasticity of VSL are sensitive to the choice of studies included in the meta-analysis. This paper systematically includes and excludes studies from the meta-analysis based on income level, offering insight regarding whether the sensitivity of a model's income elasticity to the included studies noted by Viscusi and Aldy (2003) is related in part to the income levels of the countries on which the studies are based.

Dahdah and McMahon (2008) develop an engineering rule-of-thumb approach for estimating VSL in developing countries using a regression of official VSL figures from 12 developed countries and 10 developing countries. Of the 10 developing country estimates, two are based on the WTP method, whereas eight are based on the HC method. The rule of thumb suggests a VSL value of 70 times GDP per capita with upper and lower values for sensitivity analysis of 60 and 80 times GDP per capita, respectively. An initial log-log model specification for their data yields an income elasticity of 1.125, so they opt for a simpler linear model specification because the elasticity is so close to 1.0. This rule of thumb transfer function has been widely used in developing countries by the International Road Assessment

Programme (iRAP) and by international financial institutions, such as the World Bank. Dahdah and McMahon (2008) also develop a rule of thumb for the value of a serious injury at 25% of VSL for developing countries, which is based on the VSI / VSL ratios in developed countries adjusted upwards to account for higher collision severity in developing countries. As described below, new data now exist to update the work of Dahdah and McMahon (2008) for VSL values. A primary reason for the update is that the rule of thumb was built using only 2 WTP-based values from developing countries, and the new dataset contains 123 WTP-based values from developing countries.

Hammit and Robinson (2011) raise the question of whether the income elasticity of VSL changes with income level, postulating that it makes sense for elasticity to be greater than 1.0 at low income levels and less than 1.0 at higher income levels. They review longitudinal within-country wage-risk studies, a limited sample of between-country VSL comparisons, and two within-country cross-sectional wage-risk studies employing quantile regression, finding increasing evidence that the income elasticity of VSL is greater than 1.0 at lower income levels. Hammit and Robinson (2011) indicate that a substantial degree of uncertainty exists in the VSL estimates for developing countries obtained using transfers of high-income estimates, and they suggest that more research on VSL for low-income countries is required to improve transfer functions. The elasticity evidence presented by Hammit and Robinson (2011) is based on a small sample of mostly wage-risk data. Additionally, both Bhattacharya *et al.* (2006) for India and Mahmud (2008) for Bangladesh find SP-based VSL values that are lower than the values that would

be obtained using a high-income based transfer function with an elasticity of 1.0 or nearly 1.0. These VSL values located below a linear VSL-income plot suggest convexity of the relationship (elasticity above 1.0) in the low-income range. Extensive information on SP-based VSL estimates from developing countries that became available after publication of Hammit and Robinson (2011) provides an opportunity to systematically test their elasticity findings.

The OECD (2012)<sup>2</sup> provides a rich database of VSL estimates from SP surveys in transport, health, and environmental risk contexts, including 221 estimates from developing countries. A few other developing country estimates have been published since then. The evidence presented by Hammit and Robinson (2011) for a possible income elasticity greater than 1.0 in developing countries suggests a research need to re-examine the transfer functions developed by Miller (2000) (with elasticity below 1.0) and the rule-of-thumb transfer function of Dahdah and McMahon (2008) (with elasticity = 1.0). The database provided in (OECD, 2012), along with a few additional studies published since then, presents a new opportunity to address this research need by a meta-analysis of VSL estimates from developing countries. To the authors' knowledge, this paper is the first to develop a VSL transfer function for application to transport safety engineering in

---

<sup>2</sup> (OECD 2012) compiles results from Braathen et al.(2009), Biauxque (2010), Lindhjem et al. (2011), and Lindhjem et al.(2010), which in turn draw on a large number of primary valuation studies.

developing countries that is based on VSL estimates from developing countries alone.

### **4.3. METHODOLOGY**

This section describes the methodology for the research in three subsections organized according to the objectives of the research. Overall, the methodology follows the approach used by the OECD (2012), with some departures. Section 4.3.1 describes the methodology and data sources for developing the transfer functions through meta-analytic regression models disaggregated by income level. Section 4.3.2 describes the methodology used to test for significant differences between the transfer functions obtained for different income groups. Section 4.3.3 describes the methodology for quantifying the uncertainty associated with the transfer functions.

#### **4.3.1. Data Sources and Methodology for the Development of the Transfer Function**

To obtain a developing-country VSL transfer function, the method follows five steps: (1) selecting the original database of VSL estimates; (2) disaggregating the database by income level; (3) applying quality screening to the disaggregated database of estimates; (4) establishing the characteristics of the quality-screened subset relevant to meta-analysis; and (5) selecting an appropriate model specification and a regression approach. This section outlines the methods used in each of these five steps.

#### *4.3.1.1. Selection of the original database of VSL estimates*

The original database for this meta-analysis is a comprehensive set of 862 VSL estimates from around the world. The OECD (2012) has compiled 856 of these estimates, which are available at [www.oecd.org/env/policies/vsl](http://www.oecd.org/env/policies/vsl), and we add six estimates that were published between the compilation of the OECD database and December 2013. The additional estimates are from Mongolia (Hoffman, et al., 2012), South Korea (Lee, Lim, Yang, Kim, Shin, & Shin, 2011), the United States (Viscusi, Huber, & Bell, 2013), and Sweden (Svensson, 2009). In addition to updating the OECD database with six estimates from new studies, we add a field to the database that gives the country GDP per capita from the year of the survey expressed in 2005 international dollars. The database consists only of VSL estimates based on stated preference studies. Some researchers prefer revealed preference estimates. However, the use of stated preference studies allows for the investigation of risk-context effects and the inclusion of population-representative samples (see the background discussion in Section 4.2.2).

#### *4.3.1.2. Disaggregation of the original database of VSL estimates by income levels*

Disaggregating the regression by income levels indicates whether the influence of the explanatory variables (including income) changes significantly with income level. This research represents the first income-disaggregated meta-analysis of VSL estimates in the literature. This methodology requires a threshold for income

disaggregation. The World Bank provides a widely used threshold, which is updated annually, for grouping countries according to GNI per capita. The country income groups are as follows: low income, lower middle income, upper middle income, and high income. Countries in the first three groups are commonly called *developing countries*, though the use of the term “developing country” does not imply that all countries in these groups are developing or that all countries in the high-income group have finished developing. We disaggregate the original database by comparing the survey year country GNI per capita, expressed in constant 2005 USD to the 2005 threshold between high-income and upper-middle-income countries, which is a GNI per capita of \$10,725 USD, with currency conversion by the Atlas Method.

#### *4.3.1.3. Application of quality screening to the disaggregated database of VSL estimates*

A prerequisite for a suitable transfer function is that the original studies meet an acceptable level of quality and rigor. In this study, we require that original estimates meet four criteria: (1) the survey is from a country-representative sample<sup>3</sup>; (2) the survey sample size is at least 200, and the subsample from which the estimate is derived is at least 100; (3) the original study reports the size of the risk change valued to obtain the VSL estimates; and (4) the estimates have passed an external

---

<sup>3</sup> The qualitative assessment of whether a study was ‘representative’ was made by OECD (2012); studies were excluded if they focused on a narrow group (e.g. income, students, motorcyclists, particular occupation, age, etc).

or internal scope-sensitivity test. This quality-screening approach follows one of the screening levels used in the OECD (2012) meta-analysis.<sup>4</sup> From our database of 862 estimates, 308 pass the quality screening (i.e., 123 from developing countries and 185 from high income countries). Although we adopt the quality screening criteria of the OECD, we differ in one small way when applying these criteria: the OECD rejected all estimates from a study in India (Bhattacharya, Alberini, & Cropper, 2006) because the sample of respondents was not considered representative of the country in terms of income and education. However, in reviewing this study, we found that five of the estimates were derived from a large and representative sample, whereas the remaining 13 were from non-representative sub-samples. These five estimates also met the other three quality criteria, and we retained them for our meta-analysis.

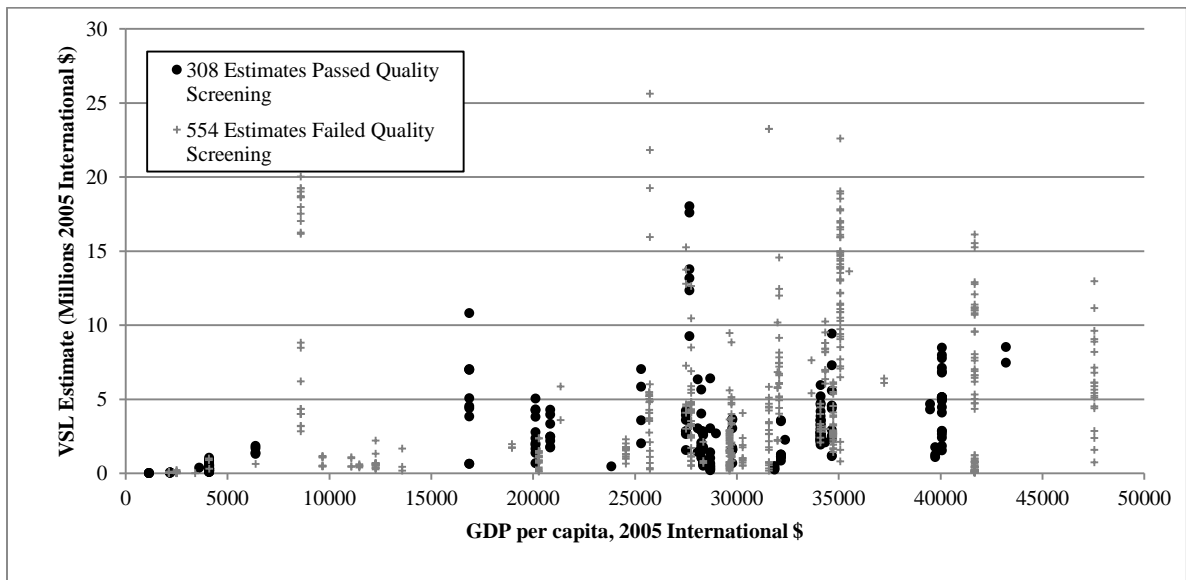
**Figure 4-1** shows the estimates that pass and those that fail the screening process plotted against GDP per capita. The Figure shows that the screening process

---

<sup>4</sup>We originally followed one of the less strict quality screening criteria from the OECD (2012) that did not require the fourth criteria of scope sensitivity. Many studies do not report scope sensitivity. We thank the reviewers for noting the importance of using only scope-sensitive estimates, which is emphasized in (Viscusi, Huber, & Bell, 2013). Scope sensitivity refers to the sensitivity of WTP answers to the size of the risk change offered in a contingent valuation survey. It can be evaluated at the individual level by offering the same person multiple risk changes (internal scope sensitivity) or at the cross-group level by offering a different risk change to multiple independent groups (external scope sensitivity) (OECD, 2012). An example of the failure of a scope sensitivity test would be if the same individual expressed the same WTP for risk changes of different magnitudes. This undermines the premise of VSL calculations ( $VSL = WTP / (\text{risk change})$ ) and suggests participant difficulty in understanding low risks.



excludes some estimates that are obvious outliers. For example, the group of exclusions includes very high estimates at just under \$10,000 GDP per capita from a study in Brazil that failed screening because the study focuses on high-income vehicle owners rather than a country-representative population. This does not mean that the Brazil study is a bad study; it simply does not suit the purposes of this research. Variability in the data at any given income level occurs partly a result of variation in other explanatory variables in addition to income (e.g., risk context). Because of this, a best-fit line in **Figure 4-1** does not necessarily represent the relationship between income and VSL in the transport risk context. Multiple linear regression accounts for the effects of these other explanatory variables.



**Figure 4-1: VSL by GDP per capita for estimates passing and failing the quality screening.** Note: To maintain reasonable vertical scale, the Figure omits 22 of the 554 estimates that fail quality screening.

*4.3.1.4. Establishment of the characteristics of the quality-screened subset relevant to meta-analysis*

Table 4-2 shows that the 123 developing country VSL estimates that pass the quality screening came from six surveys in five developing countries, with between 1 and 85 estimates per survey.

**Table 4-2: Overview of surveys and VSL estimate used for developing country transfer function**

ID <sup>a</sup>	Year	Country <sup>b</sup>	Publication	Mean VSL <sup>c</sup>	Std. Dev	N <sup>d</sup>	Scope-sensitive? <sup>e</sup>
8	2003	Thailand	(Vassanadumrondgee & Matsuoka, 2005)	1,555,256	225,888	4	Yes
34	2005	India	(Bhattacharya, Alberini, & Cropper, 2006)	41,805	10,387	5	Yes
36	2005	China	(Krupnick, Hoffman, Larsen, Peng, Tao, & Yan, 2006)	378,458	189,787	85	Yes
37	2005	China	(Krupnick, Hoffman, Larsen, Peng, Tao, & Yan, 2006)	213,545	60,789	24	Yes
38	2003	Bangladesh	(Mahmud, 2008)	3,138	707	4	Yes
-	2010	Mongolia	(Hoffman, et al., 2012)	378,275	-	1	Yes
Total						123	

Notes: <sup>a</sup>Corresponds to an identifier in the OECD dataset at [www.oecd.org/env/policies/vsl](http://www.oecd.org/env/policies/vsl). <sup>b</sup>All are classified as low- and middle-income (developing) countries according to the 2005 World Bank thresholds for 2005 GNI per capita. <sup>c</sup>The mean of the included value of statistical life (VSL) estimates from each survey, in 2005 international dollars. <sup>d</sup>The number of the included VSL estimates from each survey, used as a weight in the regression to account for each survey equally. <sup>e</sup>The study reported passing either an internal or external scope-sensitivity test. (Published paper contained minor typos in totals that are corrected here).

The five developing countries represented include countries from the low-, lower-middle-, and upper-middle-income country groups. The survey years of the estimates range from 2003 to 2010, and the mean survey VSL values range from

\$3138 to \$1.5 million. The fact that the quality screened subset contains multiple estimates per survey requires a weighting approach to avoid bias, and Section 4.3.1.5 further discusses this aspect of the method.

#### *4.3.1.5. Selection of appropriate model specification and regression approach*

The regression model specification follows (OECD, 2012), which in turn is based on standard practice in the meta-analytic literature. The model is

$$lnvsl = \beta_0 + \beta_1 lngdpcap + \sum_k \beta_k X(k) + \varepsilon, \quad (4-2)$$

where *vsl* is the value of statistical life, *gdpcap* is the PPP-adjusted GDP per capita, *X(k)* is a vector of *k* mostly binary explanatory variables describing risk context and study method (defined in Table 4.3), and the various  $\beta$ s represent the model coefficients. In this model specification, the coefficient for *lngdpcap* has the natural interpretation as the between-country income elasticity of VSL.

The model includes explanatory variables based on two revisions to those included in the fourth scope-sensitive screening model used by the OECD (OECD, 2012). Table 4-3 shows the resulting variables in the model for this research, after the following two revisions to the OECD model.

- First, the model includes the target *lnvsl* and the variable *lngdpcap* in 2005 international dollars using standard purchasing power parity exchange rates, whereas the OECD (2012) model uses purchasing power parity (PPP) exchange rates that are adjusted for actual individual consumption (AIC). AIC-adjusted values are helpful in that they reflect the individual

consumption situation better than does GDP per capita because they include non-GDP sources of consumption (e.g., foreign aid) and because they do not include the portion of GDP per capita directed away from consumption. Despite these advantages, AIC-adjusted values are available at intermittent time intervals and only for select countries through the International Comparison Program of the World Bank, whereas the more common values using standard PPP exchange rates are available for more countries and are provided on a more regular basis. The latter were selected to facilitate practical application of the transfer function.

- Second, the model removes the following explanatory variables that are constant for the developing country data set: *public* (a binary risk-context variable that is set to 1 if the survey concerned personal valuation of public risk changes as opposed to valuation of private risk changes) and *noexplain* (a methodological variable that is set to 1 if the survey included no risk explanation tools such as a 1000-square grid to help the respondents understand risks). All of the surveys in the developing country dataset valued private rather than public risk changes, and all of the surveys used risk explanation tools.

The regression approach must account for the fact that, in many cases, several VSL estimates in the database are derived from the same survey. The regression used is a weighted least-squares regression with the weights equal to the inverse of the number of estimates in a given survey to give equal weighting to each survey rather than to each estimate, following several previous studies (OECD, 2012)

(Lindhjem H. , Navrud, Braathen, & Biauxque, 2011) (Nelson & Kennedy, 2009, p. 355). The weighting can reduce the bias resulting from multiple estimates derived from one study that are potentially non-independent. Nelson and Kennedy (2009) provide a discussion on non-independence resulting from multiple estimates per study included in a meta-analysis. They find that almost 80 percent (110 of 140) of the reviewed meta-analyses used multiple estimates per study, creating a potential non-independence problem. One third of the meta-analyses (40 of 140) also implement no controls to address this potential dependency. They discuss several options to address this potential dependence. As a best option, if sufficient data are available, they recommend using only one estimate per primary study to avoid dependency problems altogether. This option is not ideal for this meta-analysis because it would result in an unacceptably low sample size. This option is also not ideal because, in several cases, the explanatory variables under investigation vary within a group of estimates in a primary study, and drawing only one estimate per primary study would reduce the ability to investigate the impact of these explanatory variables. A second option that they identify is the use of a multilevel or panel regression approach. This option is also not ideal, because some of the primary study surveys, which define the panels, have only one estimate, and could not be included in a panel approach, resulting in unacceptable information loss given the size of the data set. A third option that they identify is to apply a weighting procedure that recognizes equal contributions from each primary survey. We apply this option, which is the most suitable given our dataset and the independent variables (particularly income and risk context) that we are interested in

investigating. All regression analyses use R (R Core Team, 2013) with the robust linear modeling tools of the MASS package (Venebles & Ripley, 2002). The analysis uses robust modelling with MM estimates. Robust modelling is a group of techniques that are used to reduce the influence of unusual observations on regression results; the use of MM-estimates is a specific technique within robust modelling that offers the combined advantages of high efficiency and a high breakdown point (Yohai, 1987).

**Table 4-3: Variables included in regression models**

Variable	Type	Description
<i>lnvsl</i>	target, continuous	natural log of VSL in 2005 international dollars
<i>lngdpcap</i>	explanatory, continuous	natural log of GDP per capita in 2005 international dollars
<i>lnchrisk</i>	explanatory, continuous	natural log of the size of the risk change valued in a survey
<i>turnbull</i>	explanatory (methodological), binary	1 if the estimate is based on a turnbull lower bound estimator, 0 otherwise
<i>env</i>	explanatory (risk context), binary	1 if the estimate is based on a valuation of environmental risks, 0 otherwise
<i>traffic</i>	explanatory (risk context), binary	1 if the estimate is based on a valuation of transport related risks, 0 otherwise
<i>household</i>	explanatory (risk context), binary	1 if the estimate is based on a valuation of risks to a person's entire household, 0 otherwise
<i>cancer</i>	explanatory (risk context), binary	1 if the estimate is based on a valuation of cancer risks, 0 otherwise
<i>latent<sup>a</sup></i>	explanatory (risk context), binary	1 if the estimate is based on a valuation of risks with a long gap between risk exposure and consequence, 0 otherwise, 0 for all traffic risks

<sup>a</sup> This variable is relevant in health and environment contexts where, e.g., exposure to a carcinogenic substance can lead to a cancer death 20 years later; it is not relevant to transport risks and should be set to 0 in the transport policy context (personal communication with Ståle Navrud, Nils-Axel Braathen, and Henrik Lindhjem, authors of the OECD report, August 9, 2012).

### 4.3.2. Methodology to Test for Significant Differences Between Transfer Functions

To test for differences between transfer functions (such as different elasticity) under a null hypothesis of zero difference for each coefficient, this research uses the methodology proposed by Paternoster *et al.* (1998). This method calculates a z-score for a difference in coefficients as follows:

$$z = \frac{\beta_1 - \beta_2}{\sqrt{SE_{\beta_1}^2 + SE_{\beta_2}^2}}, \quad (4-3)$$

where  $\beta_1$  and  $\beta_2$  are the coefficients for the same variable from the two groups being tested for a difference and  $SE_{\beta_1}$  and  $SE_{\beta_2}$  are the standard errors of these coefficients. The two-tailed p-value corresponding to this z-score can be used to evaluate the statistical significance of observed differences by indicating the likelihood that the observed difference in coefficients occurred by chance alone.

### 4.3.3. Methodology to Quantify and Apply Transfer Function Uncertainty Information

To quantify transfer function uncertainty, this paper compares records in the database to the corresponding transfer function predictions, generally following the OECD (2012) method described in Section Transferring VSL Estimates Between Policy Contexts: Needs, Methods, and Uncertainty, but with some adaptations designed to offer additional insight. The first adaptation is to not use absolute values of the transfer errors because the distribution of errors is asymmetric

around zero and because taking absolute values removes information about the differences between the function over-estimates and the function under-estimates. The second adaptation is to present a set of reference percentiles of the distribution of transfer errors instead of only the mean or median absolute error value. Although the mean or median absolute error value is a good summary statistic for the dispersion of the error distribution, a set of reference percentiles of the non-absolute error distribution gives an analyst more information to apply in a sensitivity or risk analysis. Some of the limitations with mean absolute transfer error that motivate this approach are set out in Section Transferring VSL Estimates Between Policy Contexts: Needs, Methods, and Uncertainty A final adaptation is the use of the term “relative transfer error” (RTE) to emphasize that we are dealing with normalized errors (although past studies dealing with TE also have used normalized values). The adapted formula for RTE corresponding with the TE formula in Section 4.2.3 is

$$RTE = (VSL_{TF} - VSL_B) / VSL_B, \quad (4-4)$$

with the terms as described in Section 4.2.3. Using this adaptation, when the RTE distribution is analyzed, the central tendency measures no longer indicate the magnitude of a typical RTE—instead, they indicate the size and direction of any overall bias in the transfer function. The median RTE is used for this purpose because of its insensitivity to large outliers. Reference percentiles of the non-absolute RTE distribution indicate reliability of the transfer function. For example, the 25<sup>th</sup> percentile RTE gives the magnitude of a typical underestimate, and the 75<sup>th</sup> percentile RTE gives the magnitude of a typical overestimate, and half of all



RTEs lie between these two values. This also allows asymmetry in the RTE distribution to be reflected in the summary statistics, and it provides statistics with natural interpretations for practical application. This paper uses the percentile bootstrap method described by Mooney and Duval (1993) to estimate confidence intervals for the reference percentiles of the error distribution using the boot package in R (Canty & Ripley, 2013) (Davison & Hinkley, 1997).

To apply transfer function uncertainty information in a policy context, the result given by the transfer function can be adjusted correspondingly with a given ( $x^{th}$ ) percentile of the relative transfer error distribution ( $RTE_x$ ). This procedure, based on the definition of RTE, is relatively straightforward. The adjustment equation is

$$VSL_{TFadj} = (1 + RTE_x)^{-1} * VSL_{TF}, \quad (4-5)$$

where  $(1+RTE_x)^{-1}$  represents an adjustment factor linked to the  $x^{th}$  percentile of the RTE distribution,  $VSL_{TFadj}$  represents an adjusted transfer function result for sensitivity or risk analysis, and the other terms are as described earlier. There are two common types of uncertainty analyses in the engineering economic assessment portion of transport project appraisals in World Bank projects: sensitivity analysis (see, e.g., the assessment for a road safety project in Argentina (World Bank, 2010a)), and probabilistic risk analysis (see, e.g., the assessment for a highway project in Ningxia, China (World Bank, 2010b)). Sensitivity analysis checks the change in the output values corresponding to a few discrete changes in individual input values, generally using a typical or expected high and low input value. This paper proposes using the 25<sup>th</sup> and 75<sup>th</sup> percentile RTE values to

develop adjustments to the VSL transfer function results corresponding with the typical high and low values for sensitivity analysis. Probabilistic risk analysis often uses a Monte Carlo simulation that randomly varies all input values according to a triangular probability distribution to generate a simulated probability distribution for the output value. This process is described in the documentation for the *Roads Economic Decision Model (RED)* (Archondo-Callao, 2004). This process requires practical maximum and minimum values for the input parameters to define the triangle distributions. For the process to be realistic, the maximum and minimum values should result in a triangle distribution that reasonably approximates the actual distribution for the input parameter. In the case of a parameter distribution with long, flat tails, the use of actual maximums and minimums is not practical because doing so would result in a triangle distribution that over-estimates the likelihood of more extreme parameter values. Practical maximum and minimum values for generating the triangle distributions may be selected on the basis of achieving a reasonable match between the triangle distribution and the actual distribution.

#### **4.4. RESULTS**

##### **4.4.1. Transfer function regression results**

Table 4-4 shows the summary results for three transfer functions based on the weighted robust least-squares regression. Each function gives *lnvsl* as the sum of each coefficient,  $\beta$ , times the corresponding variable value for that coefficient. For example, the all-income function gives  $lnvsl = (4.387 + (.551) * lngdpcap + (-$

.534)\**Inchrisk*+ ... + (-.382)\**latent*). The  $r^2$  values in the Table show that the models have reasonably strong explanatory power. However, higher  $r^2$  values are expected in a robust regression because they indicate the amount of variability explained in the weight-transformed dataset; they should thus not be relied on as the sole criterion for a goodness-of-fit assessment (Willett & Singer, 1988). Instead, model performance is more thoroughly assessed in Section 4.4.2 with the relative transfer error distributions.

**Table 4-4: Transfer function regression results for the high-, low- and middle-, and all-income functions, with significant differences between function coefficients**

	All-income model n = 410, r <sup>2</sup> = .88			High-income model n = 258, r <sup>2</sup> = .78			Low- and middle-income model n = 152, r <sup>2</sup> = .77			High-low coefficient comparison	
	$\beta_{All_0}$	SE	p	$\beta_{HI_0}$	SE	p	$\beta_{LMI_0}$	SE	p	$\beta_{LMI_0} - \beta_{HI_0}$	p
(Intercept)	4.387	0.183	<.001	3.383	1.425	0.009	-12.941	0.383	<.001	16.324	0.000
ingdpcap	0.551	0.020	<.001	0.693	0.138	<.001	2.478	0.073	<.001	-1.785	0.000
Inchrisk	-0.534	0.011	<.001	-0.481	0.018	<.001	-0.698	0.043	<.001	0.217	0.000
turnbull	-0.557	0.077	<.001	-	-	-	-0.598	0.055	<.001	-	-
env	-0.805	0.059	<.001	-0.645	0.102	<.001	-	-	-	-	-
household	-1.419	0.054	<.001	-1.225	0.094	<.001	-	-	-	-	-
traffic	0.723	0.055	<.001	0.823	0.091	<.001	-2.377	0.128	<.001	3.200	0.000
cancer	0.390	0.049	<.001	0.470	0.080	<.001	-1.552	0.132	<.001	2.022	0.000
latent	-0.382	0.049	<.001	-0.496	0.090	<.001	-0.147	0.059	0.007	-0.350	0.001

Notes: Functions predict *lnvsl*. Variables defined in Table 4-4.  $\beta$  denotes coefficients. The income threshold is 2005 PPP-adjusted GDP per capita of \$14,000

The Table shows that the coefficients in the all-income and high-income models are fairly similar but that some significant differences exist between the coefficients for the high-income model and the low- and middle-income model. Although the between-country income elasticity (the *Ingdpcap* coefficient) for the high-income function is .693, the Table indicates a between-country income elasticity of VSL of 2.478 for the low- and middle-income country model. This is consistent with the finding of Hammit and Robinson (2011), who provide theoretical reasoning behind this finding and discuss increasing evidence in support of elasticity greater than 1 among low-income countries; it is also consistent with previous studies (Mahmud, 2008) (Bhattacharya, Alberini, & Cropper, 2006), which note that the VSL values obtained in Bangladesh and India are lower than what would be obtained under transfer from high-income countries with elasticities of 1. The difference in between-country income elasticities is significant at a p-value of less than .001.

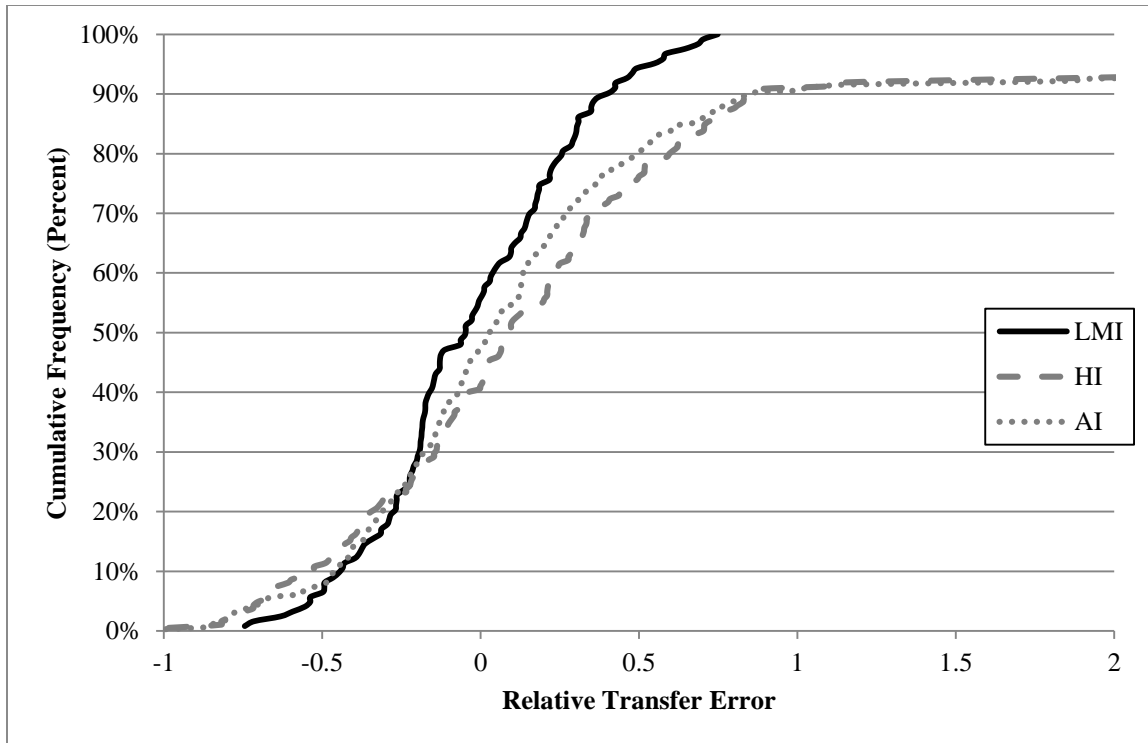
A second difference between the high-income model and the low- and middle-income model is the coefficient for the traffic-risk context parameter. If the coefficient for this parameter is positive, the WTP for traffic-related VSL is, in general, higher than the WTP for VSL in other risk contexts—people value saving a life on the road more than saving a life elsewhere. If it is negative, the opposite is true. The coefficient is significant in both models: it is positive in the high-income model (.823) and negative in the low-income model (-2.377), and the difference in the coefficient between models is significant at  $p < 0.001$ . This difference may be related to cultural factors, such as fatalistic attitudes towards traffic risks, which are explored by Nordfjærn *et al.* (2012).

A possible reason for the differences found in the low- and middle-income country model is that baseline risks and competing risks may affect VSL beyond the explanatory variables considered. Andersson and Treich (2011) present a theoretical argument for relationships between VSL and baseline risks and between VSL and competing risks, but they also note that the empirical evidence in support of these relationships is mixed. The nature of the dataset limited the ability of this meta-analysis to investigate these possible reasons in detail. For example, half of the studies in the low- and middle-income country dataset do not report baseline risk, and little information is available on competing risks.

The significantly different coefficients in the models indicate that it is more appropriate to use transfer functions based on developing country data for application in developing countries than to use transfer functions based on all- or high-income country data. Section 4.4.3 further illustrates this point for the road-safety context.

#### **4.4.2. Transfer function errors and the uncertainty analysis**

Figure 4-2 and Table 4-5 show the results of a transfer error analysis on the models by presenting the frequency distributions of relative transfer error (RTE). RTE, described in Section 4.3.3, compares the estimates in the database to what the transfer function would predict for that estimate.



**Figure 4-2: Relative transfer errors for the all-income, high-income, and low- and middle-income VSL transfer functions**

Figure 4-2 shows that the low- and middle-income country function has the narrowest RTE distribution; however, for all functions, the majority of relative transfer errors lie within +/-50%. The Figure also shows asymmetry in the RTE distributions for the high-income and all-income functions, but not in the low- and middle-income function. Table 4-5 shows the reference percentiles of the RTE distributions along with the confidence intervals for these percentiles obtained by bootstrap resampling.

**Table 4-5: Percentiles of the relative transfer error distribution with bootstrapped 95% confidence intervals for low- and middle-income, high income, and all-income transfer functions**

Percentile	RTE.LMI	RTE.HI	RTE.AI
0%	-0.745 (-0.745, -0.593)	-0.985 (-0.985, -0.798)	-0.994 (-0.994, -0.836)
5%	-0.533 (-0.624, -0.43)	-0.68 (-0.754, -0.562)	-0.675 (-0.772, -0.483)
10%	-0.434 (-0.498, -0.309)	-0.527 (-0.667, -0.413)	-0.458 (-0.532, -0.405)
15%	-0.333 (-0.442, -0.253)	-0.41 (-0.525, -0.322)	-0.37 (-0.438, -0.315)
20%	-0.268 (-0.371, -0.201)	-0.333 (-0.413, -0.203)	-0.308 (-0.363, -0.23)
25%	-0.223 (-0.29, -0.185)	-0.218 (-0.358, -0.132)	-0.229 (-0.299, -0.177)
30%	-0.191 (-0.253, -0.169)	-0.139 (-0.234, -0.078)	-0.174 (-0.229, -0.126)
35%	-0.181 (-0.212, -0.13)	-0.094 (-0.172, 0.006)	-0.127 (-0.179, -0.069)
40%	-0.156 (-0.188, -0.06)	-0.019 (-0.119, 0.064)	-0.071 (-0.129, -0.023)
45%	-0.128 (-0.176, -0.01)	0.028 (-0.066, 0.116)	-0.034 (-0.075, 0.039)
50%	-0.047 (-0.149, 0.032)	0.086 (0.01, 0.211)	0.03 (-0.036, 0.118)
55%	-0.005 (-0.125, 0.097)	0.198 (0.068, 0.276)	0.111 (0.026, 0.145)
60%	0.042 (-0.043, 0.145)	0.227 (0.121, 0.324)	0.134 (0.09, 0.212)
65%	0.114 (0.006, 0.181)	0.315 (0.213, 0.397)	0.206 (0.131, 0.275)
70%	0.161 (0.065, 0.229)	0.373 (0.284, 0.501)	0.274 (0.202, 0.365)
75%	0.202 (0.136, 0.295)	0.478 (0.333, 0.621)	0.368 (0.267, 0.483)
80%	0.258 (0.179, 0.332)	0.598 (0.445, 0.726)	0.49 (0.373, 0.603)
85%	0.308 (0.237, 0.409)	0.712 (0.568, 0.852)	0.664 (0.518, 0.789)
90%	0.4 (0.302, 0.493)	0.861 (0.707, 2.444)	0.86 (0.723, 2.276)
95%	0.543 (0.404, 0.682)	3.29 (1.072, 5.098)	4.027 (1.758, 13.528)
100%	0.749 (0.645, 0.749)	12.506 (5.888, 12.506)	67.74 (22.773, 67.74)

Notes: RTE = relative transfer error; LMI = low- and middle-income; HI = high-income; AI = all-income. Confidence intervals by percentile bootstrap method with B = 2000 resamples.

Based on Figure 4-2 and Table 4-5, the following points indicate the transfer error analysis results concerning (1) transfer function bias, (2) likely high and low values



for sensitivity analysis, and (3) practical maximum and minimum values for probabilistic risk analysis using triangle distributions.

- 1) Bias: Table 4-5 shows that there is no statistically significant bias in the low-income and all-income transfer functions and a small but a statistically significant bias in the high-income transfer function, based on the 95% CI for the median *RTE*.
- 2) Likely low and high values for sensitivity analysis: The first and third quartiles of the error distributions provide a range of likely errors when using the transfer function: half of all relative transfer errors are between these values. With the methodology explained in Section 4.3.3 and the first and third quartile *RTE* values from Table 4-5, adjustment factors  $((1 + RTE_x)^{-1})$  can be estimated to apply to the transfer function results and obtain high and low values for sensitivity analysis. The low adjustment factors are .83, .68, and .73, and the high adjustment factors are 1.29, 1.28, and 1.30 for the low- and middle-, high-, and all-income transfer functions, respectively.
- 3) Practical minimum and maximum values for probabilistic risk analysis using triangle distributions: Some forms of probabilistic risk analysis, as described in Section 4.3.3, require practical maximum and minimum values to generate a triangular distribution for an input parameter that roughly approximates the actual distribution. The high-income and all-income *RTE* distributions are positively skewed and have longer tails; the 5<sup>th</sup> and 85<sup>th</sup> percentiles of these provide values that can be used to create a reasonable

triangle distribution. The low- and middle-income RTE distribution is not seriously skewed, and its 5<sup>th</sup> and 95<sup>th</sup> percentiles can be used to create a reasonable triangle distribution. Based on Table 4-5, the adjustment factors  $((1+RTE_x)^{-1})$  to obtain practical minimums for risk analysis are .65, .58, and .60, and the adjustment factors to obtain practical maximums for risk analysis are 2.14, 3.12, and 3.07 for the low- and middle-, high-, and all-income transfer functions, respectively.

#### **4.4.3. Results applied to the transport safety policy context**

This section presents the transfer function results for the transport safety policy context. Adapting the function for this context involves setting the binary risk context and methodological variables appropriately (*turnbull* = 0; *env* = 0; *traffic* = 1; *household* = 0; *cancer* = 0; *latent* = 0), setting the risk change variable, and taking the antilog of the model defined by the coefficients in Table 4-5. Setting the risk change variable involves some ambiguity. There are two options for setting the risk change variable:

- **Option 1:** Set the risk change variable to a constant value for policy purposes. This gives a consistent VSL across road-safety projects in a given country. In this option, there is ambiguity as to what constant value should be used, beyond the fact that the value should be some proportion of the baseline risk for the group represented by the transfer function.

- **Option 2:** Set the risk change variable on a project-by-project basis according to the estimated risk change offered by the project. This would give different VSL values for different projects in a country.

For the presentation and discussion of our results, we follow Option 1. Within Option 1, we set the risk change values in each transfer function to be half of the baseline risk for the respective income group using 2011 baseline risk data from the World Health Organization (World Health Organization, 2014). The 2011 baseline health risks for road mortality in all-income, high-income, and low- and middle-income countries were 18.2, 9.0, and 19.9 per 100,000, respectively; we set the risk change values in the transfer functions at 9.1, 4.5, and 10.0 per 100,000, respectively. In an application context, a practitioner could follow Option 1 with these risk change values or Option 1 with different risk change values (e.g., a quarter of the baseline risks); alternatively, a practitioner could follow Option 2. This ambiguity in application could have been removed by not including risk change in the model, but this would have been at the expense of achieving less clarity with respect to the coefficients for other variables of interest.

With the variables set for the transport policy context, VSL becomes a function of GDP per capita. The low- and middle-income transfer function is

$$VSL_{TF,LMI} = 1.3732 \times 10^{-4} * gdp\text{cap}^{2.478}, \quad (4-6)$$

the high-income transfer function is

$$VSL_{TF,HI} = 8.2474 \times 10^3 * gdp\text{cap}^{6932}, \quad (4-7)$$

and the all-income transfer function is

$$VSL_{TF,AI} = 2.3834 \times 10^4 * gdp_{cap}^{5508}. \quad (4-8)$$

The functions are modelled for GDP per capita and VSL values expressed in 2005 international dollars, whereas a typical assessment of economic performance measures will likely require results in either US dollars or national currency values at current prices. Because the functions are non-linear with respect to income, function application using the wrong currency units can yield errors; correct function application requires attention to currency conversions and price deflators.<sup>5</sup>

**Figure 4-3** shows that the transfer functions with VSL vary by GDP per capita for the transport-policy context, illustrating the impact of modeling transfer functions for the developing-country context based on developing country data. The function modeled using high-income data provides transferred VSL values in the low- and middle-income range that are significantly higher than the functions modeled using low- and middle-income country data. The same is true for the function modeled

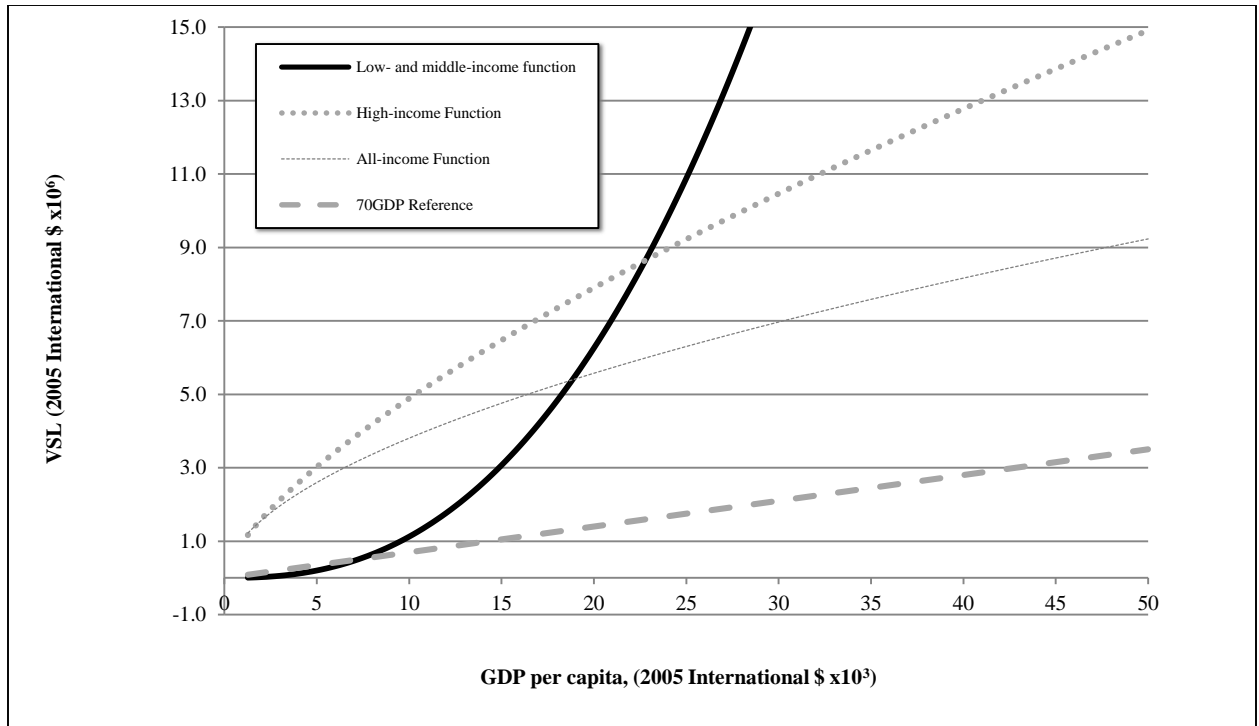
---

<sup>5</sup> The transfer function calculates  $VSL_{i,n,ID,2005}$  (country  $i$  in year  $n$  in international dollars at 2005 prices) using  $GDP\_CAP_{i,n,ID,2005}$ , which is calculated as  $GDP\_CAP_{i,n,NC,CP} * (PPP_{i,n})^{-1} * (GDP\_DEF_{US,2005} / GDP\_DEF_{US,n})$ , where  $NC$  denotes national currency,  $CP$  denotes current prices,  $PPP$  denotes the implied purchasing power parity exchange rate, and  $GDP\_DEF$  denotes the GDP deflator (to reflect changes in the price levels connected to the international dollar, which is linked to the purchasing power of the US dollar). After obtaining  $VSL_{i,n,ID,2005}$  from the transfer function,  $VSL_{i,n,NC,CP}$  (the VSL for country  $i$  in year  $n$  in national currency at current prices) may be obtained by reversing the above conversion. All of the parameters required for this conversion are available in the International Monetary Fund's *World Economic Outlook Database* (IMF, 2012).

using an aggregate of all-income data. The low- and middle-income country function crosses the all-income country function at just below \$20,000 GDP per capita and crosses the high-income function at just above \$20,000 GDP per capita; this figure of \$20,000 GDP per capita is near the upper limit of usefulness for the low- and middle-income function. The lower limit of usefulness for the low- and middle-income function is approximately \$1268 GDP per capita; there are no VSL estimates in the database below this value. Twenty-two countries, most of which are located in Sub-Saharan Africa, have estimates of the 2012 GDP per capita below this value, and at these very low values, policy makers may consider alternatives to WTP-based VSL values in the road-safety context.

The iRAP rule of thumb is a rough transfer function used by the International Road Assessment Programme and the World Bank that estimates VSL by applying a factor of 70 to GDP per capita (Dahdah & McMahon, 2008). Figure 4-3 shows that below a GDP per capita of about \$7,000, the iRAP rule of thumb gives slightly higher values than does the new low- and middle- income transfer function, whereas above \$7,000 GDP per capita, the new low- and middle-income transfer function gives significantly higher values than does the iRAP rule of thumb.

**Figure 4-3** also shows what can be interpreted as a changing economic-good nature of transport safety-risk reductions across income levels. In the lower-income range, with a between-country income elasticity of 2.478, WTP-based VSL increases more than proportionally with income; the function is convex. Among high-income countries, with a between-country income elasticity of .6932, WTP-based VSL increases less than proportionally with income; the function is concave.



**Figure 4-3: VSL by GDP per capita for the transport policy context according to new income-disaggregated benefit-transfer functions.** Note: 70GDP reference indicates the iRAP rule of thumb for calculating  $VSL = 70 * (GDP \text{ per capita})$ . Section 2.4 explains this rule of thumb (Dahdah & McMahon, 2008). In this graph, the risk-change variable in each function is set to half of the baseline road-mortality risk for the countries in the income group represented by the function.

#### 4.4.4. Results for other contexts

Whereas the main objectives of this research relate to road safety, the results can also be applied to environmental, health, and cancer risk contexts by using the coefficients from Table 4-5 and specifying appropriate values for the risk-context variables. Furthermore, the high-income function in Table 4-5 can be applied to various risk contexts in developed countries, either in the absence of appropriate local VSL estimates or as an overall appropriateness check on the reasonableness of local estimates.

#### 4.5. CONCLUSIONS

This paper presents the development of a new value of a statistical life (VSL) transfer function for application to transport-safety engineering in developing countries that is based on VSL estimates from developing countries. The transfer function estimates the value of statistical life as

$$VSL_{TF,LMI} = 1.3732 \times 10^{-4} * gdp_{cap}^{2.478} \quad (4-9)$$

where  $VSL_{TF,LMI}$  represents the value of statistical life given by the transfer function for low- and middle-income countries and  $gdp_{cap}$  represents gross domestic product (GDP) per capita expressed in 2005 international dollars. The function is applicable for countries with GDPs per capita between \$1268 and \$20,000, expressed in 2005 international dollars. Below this income range, policy makers may wish to apply alternative methods to the valuation and evaluation of transport-risk reductions. This transfer function is significantly different from a transfer function modeled on data from high-income countries, supporting this new approach, which develops transfer functions that are disaggregated by income level. In particular, the between-country income elasticity of 2.478 in developing countries, compared with that of .6932 in developed countries, is significantly different, at  $p < 0.001$ . This finding provides confirmation of initial evidence found by Hammit and Robinson (2011).

This paper also analyzes the uncertainty associated with the new transfer function using adaptations to the transfer error analysis techniques in (OECD, 2012) and other previous literature. This analysis shows that there is no statistically significant

bias in the new transfer function. Furthermore, the transfer error analysis develops adjustments to the transfer function results to support the use of VSL values in robust economic assessments during the project appraisal process.

The results and some of the database limitations suggest a need for specific further research on this subject. The number of studies providing scope-sensitive estimates from developing countries is still rather low, and all of these are from Asia. With additional scope-sensitive studies from developing countries, the following could be investigated: (1) validity of functions for very low income countries; (2) regional effects; (3) effects of further disaggregating the models within subsets of 'developing country' income levels; and (4) alternative model forms that may represent the elasticity transition across income levels change with an 'S' shaped curve that does not require model disaggregation.

To support the common practice of using one VSL per population on equity grounds (OECD, 2012), this paper does not investigate or try to predict VSL for sub-segments of a population based on factors such as income or age. While a wide literature exists on these within-country relationships, two interesting research topics to extend this literature are: (1) the impact of conditions such as a wide income gap on the estimation of an aggregate VSL for a country; and (2) appropriate policy treatments of the externalities inherent with high-income persons *creating* the bulk of mortality risk by driving and low-income persons *bearing* the bulk of mortality risk as vulnerable road users.



Taken together, these results can support the prospective performance measurement of projects impacting road safety in developing countries in a way that is practical, accounts for input uncertainty, and is amenable to application in a comprehensive assessment framework.

#### **4.6. ACKNOWLEDGEMENTS**

The authors acknowledge the support of the World Bank, the Global Road Safety Facility, and the Vanier Canada Graduate Scholarship and Michael Smith Foreign Study Supplement Programs of the Natural Sciences and Engineering Research Council of Canada. The views and findings in this paper are those of the authors and do not represent the positions of any of these agencies. The authors gratefully acknowledge guidance and comments by anonymous reviewers. The authors also acknowledge the work and generosity of the OECD in making the datasets required for this research freely and publicly available.

#### **4.7. REFERENCES**

Andersson, H, and N Treich. "The value of a statistical life." In *A Handbook of Transport Economics*, by A de Palma, R Lindsey, E Quinet, R Vickerman and eds, 396-324. Cheltenham, UK: Elgar, 2011.

Archondo-Callao, Rodrigo. *Roads Economic Decision Model Software User Guide and Case Studies (SSATP Working Paper No. 78)*. Washington, DC: The World Bank, Africa Region, 2004.

Asian Development Bank. *The cost of road traffic accidents in Malaysia*. Manila: ADB-ASEAN Regional Road Safety Program, 2005.

Bhattacharya, S, A Alberini, and M L Cropper. "The value of mortality risk reductions in Delhi, India." *The Journal of Risk and Uncertainty* 34 (2006): 21-47.

Canty, A, and B Ripley. "boot: Bootstrap R (S-plus) functions. R. Package version 1.3-9." 2013.

Cnaan, R.A., and C. Kang. "Toward valuation in social work and social services." *Research on Social Work Practice* 21, no. 4 (2011): 388-396.

Dahdah, S, and K McMahon. *The true cost of road crashes*. Hampshire, UK: International Road Assessment Program, 2008.

Dasgupta, P, A Sen, and S Marglin. *Guidlines for Project Evaluation*. New York, NY: United Nations, 1972.

Davison, A.C., and D.V. Hinkley. *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press, 1997.

Elvik, R. "Can injury prevention efforts go too far? Reflections on some possible implications of Vision Zero for road accident fatalities." *Accident Analysis and Prevention* 31, no. 3 (1999): 265-86.

Elvik, R, A Høye, T Vaa, and M Sørensen. *The Handbook of Road Safety Measures, Second Edition*. Bingley, UK: Emerald Group Publishing Ltd, 2009.

Ghani, N., and M. Faudzi. "Value of life of Malaysian motorists: Estimates from a nationwide survey." *Journal of the Eastern Asia Society for Transportation Studies* 5 (2003): 3031-3037.

Gibson, J., S. Barns, M. Cameron, S. Lim, F. Scrimgeour, and J Tressler. "The value of statistical life and the economics of landmine clearance in developing countries." *World Development* 35, no. 3 (2007): 512-531.

Giergiczny, M. "Value of a statistical life - the case of Poland." *Environmental and Resource Economics* 41 (2008): 209-221.

Goldberg, I., and J. Roosen. "Scope insensitivity in health risk reduction studies: A comparison of choice experiments and the contingent valuation method for valuing safer food." *Journal of Risk and Uncertainty* 34, no. 2 (2007): 123-144.

Grüne-Yanoff, T. "Mismeasuring the value of statistical life." *Journal of economic methodology* 16, no. 2 (2009): 109-123.

Guo, X., T. Haab, and J.K. Hammitt. "Contingent valuation and the economic value of air-pollution-related health risks in China." *3rd World Congress of Environmental and Resource Economics*. Kyoto, Japan, 2006.

Hammit, J K, and L A Robinson. "The Income Elasticity of the Value per Statistical Life: Transferring Estimates between High and Low Income Populations." *Journal of benefit-cost analysis*, 2011: 1-27.

Hammitt, J.K., and Y. Zhou. "The economic value of air-pollution-related health risks in China: A contingent valuation study." *Environmental and Resource Economics* 33 (2006): 399-423.

Hauer, E. "Computing what the public wants: some issues in road safety cost-benefit analysis." *Accident Analysis and Prevention* 43, no. 1 (2011): 151-164.

Hoffman, S, et al. "The willingness to pay for mortality risk reductions in Mongolia." *Resource and Energy Economics*, 2012: 493-513.

IMF. *World Economic Outlook Database*. April 2012. <http://www.imf.org/external/pubs/ft/weo/2012/01/weodata/index.aspx> (accessed August 8, 2012).

Kochi, I, B Hubbell, and R Kramer. "An empirical Bayes approach to combining and comparing estimates of the value of a statistical life for environmental policy analysis." *Environmental and Resource Economics* 34, no. 3 (2006): 385-406.

Kristofersson, D, and S Navrud. "Can use and non-use values be transferred across countries?" In *Environmental value transfer: Issues and methods*, by S Navrud and R Ready (ed). Dordrecht, The Netherlands: Springer, 2007.

Kristofersson, D, and S Navrud. "Validity tests of benefit transfer - are we performing the wrong tests?" *Environmental and resource economics* 30 (2005): 279-286.

Krupnick, A, S. Hoffman, B. Larsen, X. Peng, R Tao, and C. Yan. *Mortality risk valuation in Shanghai and Chongqing, China*. Washington, DC: Resources for the Future, 2006.

Lee, Y. J., Y. W Lim, J.Y. Yang, C.S. Kim, Y. C. Shin, and D.C. Shin. "Evaluating the PM damage cost due to urban air pollution and vehicle emissions in Seoul, Korea." *Journal of Environmental Management*, 2011: 603-609.

Lindhjem, H, S Navrud, N.A. Braathen, and V Biousque. "Valuing mortality risk reductions from environmental, transport and health policies: a global meta-analysis of stated preference studies." *Risk Analysis* 31, no. 9 (2011): 1381-1407.

Lindhjem, Henrik, and Stale Navrud. "How reliable are meta0analyses for international benefit transfers?" *Ecological Economics*, 2008: 425-435.

Mahmud, Minhaj. "On the contingent valuation of mortality risk reduction in developing countries." *Applied Economics*, 2008: 171-181.

Manheim, M. *Fundamentals of Transportation Systems Analysis, Volume 1: Basic Concepts*. Cambridge, MA: MIT Press, 1979.

McConnel, K. "Hedonic Wage Analysis." In *Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models (The Economics of Non-Market Goods and Resources)*, 189-242. Springer, 2006.

Miller, T.R. "Variations between countries in values of statistical life." *Journal of Transport Economics and Policy* 34, no. 2 (2000): 169-188.

Mooney, C Z, and R D Duval. *Bootstrapping: a nonparametric approach to statistical inference (Sage University Paper Series on Quantitative Applications in the Social Sciences series no. 07-095)*. Newbury Park, CA: Sage, 1993.

Mrozek, J.R., and L.O. Taylor. "What determines the value of life? A meta analysis." *Journal of Policy Analysis and Management* 21, no. 2 (2002): 253-270.

Navrud, S., and R. Ready. "Review of methods for value transfer." In *Environmental value transfer: Issues and methods*, by S. Navrud and R. Ready (ed.). Dordrecht, The Netherlands: Springer, 2007.

Nelson, J P, and P E Kennedy. "The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment." *Environmental and Resource Economics*, 2009: 345-377.

Nordfjærn, T, S Jørgensen, and T Rundmo. "Cultural and socio-demographic predictors of car accident involvement in Norway, Ghana, Tanzania and Uganda." *Safety Science* 50, no. 9 (2012): 1862-1872.

OECD. *Mortality Risk Valuation in Environment, Health and Transport Policies*. Paris: OECD Publishing, 2012.

Paternoster, R, R Brame, P Mazerolle, and A Piquero. "Using the correct statistical test for the equality of regression coefficients." *Criminology* 36, no. 4 (1998): 859-866.

R Core Team. *R: A language and environment for statistical computing*. Vienna, Australia: R Foundation for Statistical Computing, 2013.

Rosenberger, R.S., and J.B. Loomis. "Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation." *Water Resources Research*, 2000: 1097-1107.

Rosencrantz, H, K Edvarsson, and S.O. Hansson. "Vision Zero - Is it irrational?" *Transportation Research Part A: Policy and Practice*, 2007: 559-567.

Ruser, J, and R. Butler. "The economics of occupational safety and health." *Foundations and trends in microeconomics* 5, no. 5 (2009): 301-354.

Svensson, M. "The value of statistical life in Sweden: Estimates from two studies using the "Certainty Approach" calibration." *Accident Analysis and Prevention*, 2009: 430-437.

Vassanadumrongdee, S., and S. Matsuoka. "Risk perceptions and value of a statistical life for air pollution and traffic accidents: Evidence from Bangkok, Thailand." *Journal of Risk and Uncertainty* 30 (2005): 261-287.

Venebles, W.N., and B.D. Ripley. *Modern Applied Statistics with S. Fourth Edition*. New York: Springer, 2002.

Viscusi, W.K., and J Aldy. "The value of a statistical life: a critical review of market estimates throughout the world." *Journal of Risk and Uncertainty* 27, no. 1 (2003): 5-76.

Viscusi, W.K., J Huber, and J. Bell. "Assessing whether there is a cancer premium for the value of a statistical life." *Health Economics*, 2013.

WHO. "Second Global Status Report on Road Safety - Report Flyer." *World Health Organization*. 2012.

[http://www.who.int/violence\\_injury\\_prevention/global\\_status\\_report/flyer\\_en.pdf](http://www.who.int/violence_injury_prevention/global_status_report/flyer_en.pdf)  
(accessed 11 7, 2012).

Willett, J.B, and J.D. Singer. "Another Cautionary Note about R2: Its Use in Weighted Least-Squares Regression Analysis." *The American Statistician* 42, no. 3 (1988): 236-238.

World Bank and WHO. *World report on road traffic injury prevention*. Geneva: World Health Organization, 2004.

World Bank. *Project appraisal document on a proposed loan in the amount of US\$250 million to the People's Republic of China for a Ningxia Highway Project*. Washington, DC: The World Bank, 2010b.

World Bank. *Project appraisal document on a proposed loan in the amount of US\$30 million to the Argentine Republic for a road safety project in support of the first phase (APL-1) of a national road safety program*. Washington, DC: World Bank, 2010a.

—. *Safe, Clean and Affordable Transport for Development: The World Bank Group's Transport Business Strategy for 2008-2012*. Washington, DC: World Bank Group, 2008.



World Health Organization. "WHO Health Statistics and Health Information Systems." *Regional estimates for 2000-2011, Cause Specific Mortality, World Bank Income Groups.* 2014.  
[http://www.who.int/entity/healthinfo/global\\_burden\\_disease/GHE\\_DthWBInc\\_2000\\_2011.xls?ua=1](http://www.who.int/entity/healthinfo/global_burden_disease/GHE_DthWBInc_2000_2011.xls?ua=1) (accessed 02 21, 2014).

Yohai, V.J. "High breakdown-point and high efficiency robust estimates for regression." *The Annals of Statistics* 15, no. 20 (1987): 642-656.

## **5. CONCLUSION**

### **5.1. PERFORMANCE MEASUREMENT AND RISK ANALYSIS AS FOUNDATIONS FOR TRANSPORT AGENCY MANAGEMENT**

Organizations tasked with fulfilling a mandate – like transport agencies – can often benefit by expressing that mandate in a series of performance measures which become a focal point for management tasks such as: making decisions about resource allocation, design, operations, regulation, or maintenance; creating accountability structures; and creating agency focus. The ability of performance measures to serve as such a focal point depends on the certainty levels in the measures or forecasts of these measures. Where uncertainty exists, an understanding of that uncertainty helps provide proper context for professionals relying on performance measures as a focal point for management tasks. One way to understand uncertainty in performance measures is to conduct a risk analysis.

In this research, risk analysis refers to clarifying uncertainty for decision-makers. While there are many techniques for risk analysis, this research deals specifically with probabilistic risk analysis. This type of risk analysis generates estimates of probability distributions of future performance measure values, usually with Monte Carlo simulation methods. These methods require information on input uncertainty, which can be either subjective or objective information. Objective information on input uncertainty provides a stronger foundation for risk analysis, leading to more confident clarification of performance measure forecast uncertainty for decision-makers. In road safety, the ability to conduct a risk analysis

of performance measure forecasts has been constrained by the lack of information on input uncertainty.

## **5.2. CONTRIBUTIONS TO KNOWLEDGE**

The research makes methodological, empirical, and ancillary contributions to knowledge in the fields of road safety engineering, performance measurement, and risk analysis. The methodological contribution to knowledge is the design and demonstration of methods to characterize input uncertainty for application to probabilistic risk analysis of performance measure forecasts. The method is transferable across jurisdictions, across disciplines within transportation engineering and planning, and to other socio-technical application contexts where performance models use uncertain inputs to forecast uncertain outcomes. In all cases, method transferability requires the ability to establish a sample of inputs for which ground truth is available.

The empirical contributions consist of the first quantifications of uncertainty for three categories of inputs to road safety performance measure forecasts in specific contexts, based on the results of the three experimental demonstrations of the methodology. These are: (1) the uncertainty characteristics of annual pedestrian crossing risk exposure estimates when these estimates are derived from expanding short-term counts with external temporal information at a downtown site in Winnipeg, Manitoba; (2) the uncertainty characteristics of vehicle risk exposure estimates for highways in Manitoba when these estimates are derived from expanding short-term counts using temporal information from the individual

permanent counter (IPC) method; and (3) the uncertainty characteristics of value of statistical life estimates when these estimates are obtained using an income-disaggregated meta-analysis transfer function.

The empirical results for uncertainty in pedestrian crossing risk exposure estimates are based on a database of pedestrian flows from video review covering 12 months and including over 350,000 pedestrian observations. Exposure estimates developed by expanding counts with local vehicle factors have the lowest errors (mean: -2%; median: -3%, standard deviation: 33%; 90 percent of errors between -53% and 50%). Exposure estimates based on external composite pedestrian patterns have higher errors (mean: 27%; median: 9%; standard deviation: 73%; 90 percent of errors between -62% and 170%).

The empirical results for uncertainty in vehicle risk exposure estimates are based on a set of almost 2 million short-term counts at 69 sites in Manitoba, Canada. The results show that exposure estimates developed by expanding short-term counts using the individual permanent counter method have mean absolute transfer error of 6.7% (varies by traffic pattern group from 5% to 10.5%), and that over 90% of errors lie between -20% and +30%.

The empirical results for uncertainty in value of statistical life (VSL) estimates are based on testing a database of 123 scope-sensitive VSL estimates from developing countries against results of an income-disaggregated VSL transfer function. The results show that the income-disaggregated low- and middle-income function is unbiased, with a median relative transfer error of -.05 (95% CI: -.15 to

.03), a 25<sup>th</sup> percentile relative transfer error of -.22 (95% CI: -.29 to -.19), and a 75<sup>th</sup> percentile relative transfer error of .20 (95% CI: .14 to .30).

The internal validity of the empirical results (the robustness of the results within the context of the studies) is tested and described using non-parametric techniques for confidence interval estimation (in the case of expanded vehicle counts this step is omitted because the sample size of 2 million error estimates makes bootstrapping the results unnecessary for assessing internal validity). The external validity of the empirical contributions – whether similar uncertainty characteristics hold for the same inputs in different contexts – is not assessed and is outside the scope of this research. The only way to determine external validity of the results – and the conditions under which such validity holds – is to repeat the experiments across contexts and compare the results of the error distributions.

The main ancillary contribution to knowledge – one that is not directly linked to the connected theme of the papers – consists of substantial advances in the practice of VSL estimation by transfer function for developing countries. VSL is an important input to economic performance measures in road safety engineering. Before testing uncertainty in VSL estimates, this research develops the first income-disaggregated benefit transfer function for VSL in road safety in developing countries. Previous transfer functions for application in developing countries were based largely on high-income country data that hid the unique relationship between VSL and its influencing factors in developing countries. For example, the transfer function used before this research to guide the road safety evaluation of \$4 to \$7 billion in annual transport lending by the World Bank was based on meta-

analysis of a dataset that contained 2 data points from developing countries and 25 data points from high-income countries (Dahdah & McMahon, 2008). The function developed in this research, to be used by the World Bank going forward, uses a dataset with 123 data points from developing countries and no data points from high-income countries. In addition to providing the first income-disaggregated transfer function, the research contributes the first statistically significant empirical evidence for two phenomena related to willingness to pay for road safety mortality risk reduction that have previously been postulated but never quantitatively proven. The first phenomenon is that the income-elasticity of VSL is higher for developing countries than it is for high-income countries – Hammit and Robinson (2011) outline the postulated case for this. The second phenomenon is that the relative valuation of road mortality risk reduction vis-à-vis other categories of mortality risk reduction is much lower in developing countries than the same relative valuation in high-income countries - Nordfjærn et al. (2012) outline the postulated case for this phenomenon. The income-disaggregated transfer function development approach allows quantitative testing of these phenomena by comparison of function coefficients between the developing country and high-income country functions.

These contributions to knowledge can serve to advance and improve the practice of risk analysis of road safety performance forecasts, which in turn can generate a clearer understanding of key issues for investment decision-makers. While advancing the state of academic research, these contributions to knowledge also have direct relevance to engineering practitioners in the following five ways:

- (1) A common comment by practicing engineers in road safety when considering use of advanced tools starts like this: “but our data”. The comment reflects a sense that uncertainty in the input data can limit the usefulness of advanced. For these hesitant practitioners, the research shows that the advanced tools can still be used with imperfect input data by applying quantitative approaches to deal with the input uncertainty.
- (2) A practicing engineer can use these results to perform quick sensitivity checks on the results of their analysis.
- (3) A practicing engineer who is using advanced forecasting methods can employ a simplified version of these studies to quantify the uncertainty in their own inputs, thereby gaining a better understanding of the limitations inherent in their results.
- (4) A practicing engineer can anticipate that future forecasting tools may increasingly incorporate risk-analysis approaches that require understanding of input uncertainty.
- (5) Given the fact that some input uncertainties revealed in this research are quite large, a practicing engineer at a highway agency procuring a safety analysis through a consultant may stipulate that the consultant explicitly address analysis input uncertainty and the resulting impact on forecasts.

### **5.3. RECOMMENDATIONS FOR FUTURE RESEARCH**

Several areas connected to the theme of this research but outside of its scope could benefit from further investigation. More inputs to road safety performance forecasts could be tested (for example, the input parameters of microsimulation models used to create some safety performance forecasts are subject to uncertainty). The external validity of the empirical contributions can be evaluated by conducting similar experiments in different contexts (for example, repeating the pedestrian crossing volume experiment at locations in other Canadian cities). Additional methods for estimating inputs can be tested (for example, expanding short-term vehicle counts with the group factor method). While this research focused on establishing the degree of input uncertainty relevant to risk analysis, further research could investigate the root causes of this uncertainty and mechanisms to reduce it for gaining more confident performance measure forecasts (for example, the methods for arranging sites into traffic pattern groups can impact the uncertainty of AADT estimates). In the case of VSL estimates, a transfer function leading to lower VSL uncertainty may be obtained by a new meta-analysis after an increased number of original studies are conducted in more low-income countries. Hertz (1964) noted that our performance forecast capabilities advanced beyond our probabilistic risk analysis capabilities, potentially leading to false confidence in performance forecasts. Since then, probabilistic risk analysis techniques such as Monte Carlo simulation have advanced and become widely understood. However, the development of these techniques has in turn outpaced the growth of the evidence base on input uncertainty that the techniques require.



In the area of road safety engineering, continued population of this evidence base fills a knowledge gap that enables risk analysis of performance forecasts, which in turn enables stronger decision-making.

## **6. REFERENCES**

AASHTO. (2010). *Highway Safety Manual 1st Edition*. Washington, DC: American Association of State Highway and Transportation Officials (AASHTO).

AASHTO. (2011). *Transportation Asset Management Guide: A Focus on Implementation*. Washington DC: American Association of State Highway and Transportation Officials.

Abbott, E. E., Cantalupo, J., & Dixon, L. B. (1998). Performance Measures: Linking Outputs and Outcomes to Achieve Goals. *Transportation Research Record: Journal of the Transportation Research Board*, 1617, 90-95.

Albright, D. (1987). A Quick Cluster Control Method: Permanent Control Station Cluster Analysis in Average Daily Traffic Calculations. *Transportation Research Record*, 1134, 57-64.

Andersson, H., & Treich, N. (2011). The value of a statistical life. In A. de Palma, R. Lindsey, E. Quinet, R. Vickerman, & eds, *A Handbook of Transport Economics* (pp. 396-324). Cheltenham, UK: Elgar.

Archondo-Callao, R. (2004). *Roads Economic Decision Model Software User Guide and Case Studies (SSATP Working Paper No. 78)*. Washington, DC: The World Bank, Africa Region.

Aultman-Hall, L., Lane, D., & Lambert, R. R. (2009). Assessing Impact of Weather and Season on Pedestrian Traffic Volumes. *Transportation Research Record, No. 2140*, 35-43.

Baird, M., & Stammer Jr., R. (2000). Measuring the Performance of State Transportation Agencies: Three Perspectives. *Transportation Research Record: Journal of the Transportation Research Board*, 1729, 26-34.

Ball, C., & Dunn, P. (1997). Performance measurement in public civil works agencies. *Engineering Management Journal*, 9(4), 19-23.

Bhattacharya, S., Alberini, A., & Cropper, M. L. (2006). The value of mortality risk reductions in Delhi, India. *The Journal of Risk and Uncertainty*, 34, 21-47.

Bishop-Edkins, C., & Nethercut, C. (1988). Initiating the Strategic Planning Process at NJ Transit. *Transportation Research Record: Journal of the Transportation Research Board*(1156), 81-87.

Bodle, R. (1966). *Evaluation of Rural Coverage Count Duration for Estimating Annual Average Daily Traffic*. Washington, DC: U.S. Department of Commerce Bureau of Public Roads.

Börjesson, M., & Eliasson, J. (2011). On the use of "average delay" as a measure of train reliability. *Transportation Research Part A*, 45(3), 171-184.

Bremmer, D., & Bryan, J. (2008a). Bridging the Gap Between Agencies and Citizens: Performance Journalism as a Practical Solution to Communicate

Performance Measures and Results. *Transportation Research Record: Journal of the Transportation Research Board*(2046), 20-29.

Bremmer, D., & Bryan, J. (2008b). Making the Case for Funding Using Performance Management: Experience of the Washington State Department of Transportation. *Transportation Research Record: Journal of the Transportation Research Board*(2079), 146-153.

Bremmer, D., Cotton, C., & Hamilton, B. (2005). Emerging Performance Measurement Responses to Changing Political Pressures at State Departments of Transportation: Practitioner's Perspective. *Transportation Research Record: Journal of the Transportation Research Board*(1924), 175-183.

Bu, F., Greene-Roesel, R., Diogenes, M. C., & Ragland, D. (2007). *Estimating Pedestrian Accident Exposure: Automated Pedestrian Counting Devices Report*. California: Safe Transportation Research & Education Center, Institute of Transportation Studies, UC Berkeley.

Cambridge Systematics, Inc. & High Street Consulting Group. (2010). *Transportation Performance Management: Insight from Practitioners*. TRB, Washington, DC.

Campbell, J. F. (1995). Using Small Trucks to Circumvent Large Truck Restrictions: Impacts on Truck Emissions and Performance Measures. *Transportation Research Part A*, 29A(6), 445-458.

- Canning, P., Hellowell, E., Hughes, S., Gatersleben, B., & Fairhead, C. (2010). "Devolution' of transport powers to Local Government: Impacts of the 2004 Traffic Management Act in England. *Transport Policy*, 17(2), 64-71.
- Canty, A., & Ripley, B. (2013). boot: Bootstrap R (S-plus) functions. R. Package version 1.3-9.
- Chen, A., & Kasikitwiwat. (2011). Modeling capacity flexibility of transportation networks. *Transportation Research Part A*, 45(2), 105-117.
- Chen, S. (1981). *Improved Methods for State-Wide Vehicle Counting*. West Lafayette, Indiana: Purdue University Joint Highway Research Project.
- Cnaan, R., & Kang, C. (2011). Toward valuation in social work and social services. *Research on Social Work Practice*, 21(4), 388-396.
- Cottrell, W., & Pal, D. (2003). Evaluation of Pedestrian Data Needs and Collection Efforts. *Transportation Research Record, Journal of the Transportation Research Board* 1828, 12-19.
- Curtis, J., D'Angelo, D., Hallowell, M., Henkel, T., & Molenaar, K. (2012). Enterprise risk management for transportation agencies. *Transportation Research Record: Journal of the Transportation Research Board*, 2271, 57-65.
- Dahdah, S., & McMahon, K. (2008). *The true cost of road crashes*. Hampshire, UK: International Road Assessment Program.

Dasgupta, P., Sen, A., & Marglin, S. (1972). *Guidelines for Project Evaluation*. New York, NY: United Nations.

Davison, A., & Hinkley, D. (1997). *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press.

DoT. (1996). *National Cycling Strategy*. London: Department of Transport.

Effron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 54-75.

El-Bayoumy, K., & Sayed, T. (2010). Safety performance functions with measurement errors in traffic volume. *Safety Science*, 48, 1339-1344.

Elvik, R. (1999). Can injury prevention efforts go too far? Reflections on some possible implications of Vision Zero for road accident fatalities. *Accident Analysis and Prevention*, 31(3), 265-86.

Elvik, R., Høyee, A., Vaa, T., & Sørensen, M. (2009). *The Handbook of Road Safety Measures, Second Edition*. Bingley, UK: Emerald Group Publishing Ltd.

Esawey, M., Lim, C., & Sayed, T. M. (2013). Development of daily adjustment factors for bicycle traffic. *Journal of Transportation Engineering*, 139(8), 859-871.

Etmanczyk, J. (1995). Strategic planning, total quality and performance measurement: A quality director's view. *Transportation Research Record: Journal of the Transportation Research Board*, 1498, 65-70.

FHWA. (2012, July 17). *MAP-21 - Moving Ahead for Progress in the 21st Century: A Summary of Highway Provisions*. Retrieved September 18, 2012, from <http://www.fhwa.dot.gov/map21/summaryinfo.cfm>

FHWA. (2013). *Crash Modification Factors Clearinghouse*. Retrieved 01 11, 2013, from <http://www.cmfclearinghouse.org/>

FHWA. (2013). *Traffic Monitoring Guide*. Washington, DC: United States Federal Highways Administration.

Fitzpatrick, K., & Park, E. S. (2009). Safety Effectiveness of HAWK Pedestrian Treatment. *Transportation Research Record, Journal of the Transportation Research Board, No. 2140*, 214-223.

Foote, P. (2004). Making Buses Better in Chicago: Strategic Implementation of Customer-Derived Performance Measures from 1995 to 2001. *Transportation Research Record: Journal of the Transportation Research Board*(1884).

Francis, G., Fry, J., & Humphreys, I. (2002). International Survey of Performance Measures in Airports. *Transportation Research Record: Journal of the Transportation Research Board*, 1788, 101-106.

Gadda, S., Kockelman, K., & Magoon, A. (2007). Estimates of AADT: Quantifying the Uncertainty. *World Conference on Transportation Research* (pp. 1-17). Berkley, CA: WCTR.

Galway, L.A. (2007). Subjective Probability Distribution Elicitation in Cost Risk Analysis. RAND Corporation: Santa Monica, CA.

Gargett, A., & Wallis, I. (1989). Transit service performance assessment in Adelaide. *Transport Reviews*, 9(1), 75-92.

Ghani, N., & Faudzi, M. (2003). Value of life of Malaysian motorists: Estimates from a nationwide survey. *Journal of the Eastern Asia Society for Transportation Studies*, 5, 3031-3037.

Gibson, J., Barns, S., Cameron, M., Lim, S., Scrimgeour, F., & Tressler, J. (2007). The value of statistical life and the economics of landmine clearance in developing countries. *World Development*, 35(3), 512-531.

Giergiczny, M. (2008). Value of a statistical life - the case of Poland. *Environmental and Resource Economics*, 41, 209-221.

Gifford, J., & Carlisle, W. (2004). Virginia Department of Transportation's Dashboard Performance Measurement and Reporting System: Going the Full Monty. *Transportation Research Record: Journal of the Transportation Research Board*(1895), 207-219.

Gihring, C. K., & Greene, W. (2000). Washington State Ferries: Performance Measures and Information Support. *Transportation Research Record*, 1704, 93-99.

Giuliano, G. (2007). The Changing Landscape of Transportation Decision Making. *Transportation Research Record: Journal of the Transportation Research Board*, No 2036, 5-12.

Goldberg, I., & Roosen, J. (2007). Scope insensitivity in health risk reduction studies: A comparison of choice experiments and the contingent valuation method for valuing safer food. *Journal of Risk and Uncertainty*, 34(2), 123-144.

Goldman, T., & Deakin, E. (2000). Regionalism through partnerships? Metropolitan Planning Since ISTEA. *Berkeley Planning Journal*, 14(1), 46-75.

Government of the United Kingdom. (2005). How the National Road Traffic Estimates are made. *Workshop on statistics on the volume of road traffic (vehicle-kilometres)* (pp. 1-6). Copenhagen: United Nations Economic Commission for Europe.

Granato, S. (1998). The Impact of Factoring Traffic Counts for Daily and Monthly Variation in Reducing Sample Counting Error. *Proceedings of the Crossroads 2000 Conference* (pp. 122-125). Ames, Iowa: Iowa State University Center for Transportation Research and Education.

Grüne-Yanoff, T. (2009). Mismeasuring the value of statistical life. *Journal of economic methodology*, 16(2), 109-123.

Guerre, J., & Evans, J. (2009). Applying System-Level Performance Measures and Targets in the Detroit, Michigan, Metropolitan Planning Process. *Transportation Research Record: Journal of the Transportation Research Board*, 2119, 27-35.



Guo, X., Haab, T., & Hammitt, J. (2006). Contingent valuation and the economic value of air-pollution-related health risks in China. *3rd World Congress of Environmental and Resource Economics*. Kyoto, Japan.

Hammit, J. K., & Robinson, L. A. (2011). The Income Elasticity of the Value per Statistical Life: Transferring Estimates between High and Low Income Populations. *Journal of benefit-cost analysis*, 1-27.

Hammitt, J., & Zhou, Y. (2006). The economic value of air-pollution-related health risks in China: A contingent valuation study. *Environmental and Resource Economics*, 33, 399-423.

Handy, S. (2008). Regional transportation planning in the US: An examination of changes in technical aspects of the planning process in response to changing goals. *Transport Policy*, 15(2), 113-126.

Hauer, E. (1997). *Observational Before-After Studies in Road Safety*. Oxford: Elsevier Science Ltd.

Hauer, E. (2011). Computing what the public wants: some issues in road safety cost-benefit analysis. *Accident Analysis and Prevention*, 43(1), 151-164.

Hauer, E. (2014). *The art of regression modeling in road safety* (January 2014 ed.). Toronto, Canada: Ezra Hauer.

Hauer, E., Kononov, J., Allery, B., & Griffith, M. (2002). Screening the Road Network for Sites with Promise. *Transportation Research Record: Journal of the Transportation Research Board*, 1784, 27-32.

Henning, T. F., Muruvan, S., Feng, W. A., & Dunn, R. (2011). The development of a benchmarking tool for monitoring progress towards sustainable transportation in New Zealand. *Transport Policy*, 18, 480-488.

Hernandez, H., Baumgartner, T., Poapst, R., MacAngus, J., Milligan, C., Moshiri, M., et al. (2011). *Traffic on Manitoba Highways 2010*. Winnipeg, MB: Manitoba Highway Traffic Information System.

Hertz, D. (1964). Risk Analysis in Capital Investment. *Harvard Business Review*, 95-106.

Hickman, R., & Banister, D. (2007). Looking over the horizon: Transport and reduced CO2 emissions in the UK by 2030. *Transport Policy*, 14(5), 377-387.

Hoffman, S., Qin, P., Krupnick, A., Badrakh, B., Batbaatar, S., Altangerel, E., et al. (2012). The willingness to pay for mortality risk reductions in Mongolia. *Resource and Energy Economics*, 493-513.

Howard, M. (1988). Successfully Establishing a Strategic Planning Process. *Transportation Research Record: Journal of the Transportation Research Board*(1156), 73-80.

Hu, P. S., Wright, T., & Esteve, T. (1998). Traffic Count Estimates for Short-Term Traffic Monitoring: Simulation Study. *Transportation Research Record*, 26-34.

Hudson, J., Qu, T.-B., & Turner, S. (2010). *Forecasting Bicycle and Pedestrian Usage and Research Data Collection Equipment*. College Station, TX: Texas Transportation Institute.

IMF. (2012, April). *World Economic Outlook Database*. Retrieved August 8, 2012, from <http://www.imf.org/external/pubs/ft/weo/2012/01/weodata/index.aspx>

Ismail, K., & Sayed, T. (2012). Risk-optimal highway design: Methodology and case studies. *Safety Science*, 50, 1513-1521.

ISO. (2009). ISO Guide 73:2009: Risk management - Vocabulary. International Organization for Standardization. Geneva. Retrieved December 9<sup>th</sup>, 2014 from <https://www.iso.org/obp/ui/#iso:std:iso:guide:73:ed-1:v1:en>.

Jiang, Z., McCord, M. R., & Goel, P. K. (2006). Improved AADT Estimation by Combining Information in Image- and Ground-Based Traffic Data. *Journal of Transportation Engineering*, 523-530.

Karlaftis, M., & Tsamboulas, D. (2012). Efficiency measurement in public transport: Are findings specification sensitive? *Transport Research Part A*, 46(2), 392-402.

Kassoff, H. (2001). Implementing Performance Measurement in Transportation Agencies. *Conference Proceedings 26: Performance Measures to Improve*

*Transportation Systems and Agency Operations* (pp. 47-58). Irvine, CA: Transportation Research Board.

Knoop, V., Snelder, M., van Zuylen, H., & Hoogendoorn, S. (2012). Link-level vulnerability indicators for real-world networks. *Transportation Research Part A*, 46(5), 843-854.

Kochi, I., Hubbell, B., & Kramer, R. (2006). An empirical Bayes approach to combining and comparing estimates of the value of a statistical life for environmental policy analysis. *Environmental and Resource Economics*, 34(3), 385-406.

Kristofersson, D., & Navrud, S. (2005). Validity tests of benefit transfer - are we performing the wrong tests? *Environmental and resource economics*, 30, 279-286.

Kristofersson, D., & Navrud, S. (2007). Can use and non-use values be transferred across countries? In S. Navrud, & R. Ready (ed), *Environmental value transfer: Issues and methods*. Dordrecht, The Netherlands: Springer.

Krupnick, A., Hoffman, S., Larsen, B., Peng, X., Tao, R., & Yan, C. (2006). *Mortality risk valuation in Shanghai and Chongqing, China*. Washington, DC: Resources for the Future.

Lee, D. (1989). Transit cost and performance measurement. *Transport Reviews*, 9(2), 147-170.

Lee, Y. J., Lim, Y. W., Yang, J., Kim, C., Shin, Y. C., & Shin, D. (2011). Evaluating the PM damage cost due to urban air pollution and vehicle emissions in Seoul, Korea. *Journal of Environmental Management*, 603-609.

Li, M.-T., Zhao, F., & Chow, L.-F. (2006). Assignment of Seasonal Factor Categories to Urban Coverage Count Stations Using a Fuzzy Decision Tree. *Journal of Transportation Engineering*, 654-662.

Lindhjem, H., & Navrud, S. (2008). How reliable are meta-analyses for international benefit transfers? *Ecological Economics*, 425-435.

Lindhjem, H., Navrud, S., Braathen, N., & Biaisque, V. (2011). Valuing mortality risk reductions from environmental, transport and health policies: a global meta-analysis of stated preference studies. *Risk Analysis*, 31(9), 1381-1407.

Lucas, B. (1996). *Design, Development, and Implementation of the Manitoba Highway Traffic Information System (Master's Thesis)*. Winnipeg: University of Manitoba.

Maher, M. J., & Summersgill, I. (1996). A comprehensive methodology for the fitting of predictive accident models. *Accident Analysis and Prevention*, 281-296.

Mahmud, M. (2008). On the contingent valuation of mortality risk reduction in developing countries. *Applied Economics*, 171-181.

Manheim, M. (1979). *Fundamentals of Transportation Systems Analysis Volume 1: Basic Concepts* (Second Printing ed.). Cambridge, MA: The MIT Press.

Manzetti, L. (2003). Political Manipulations and Market Reform Failures. *World Politics*, 55, 315-360.

Marsden, G. (2009). Impacts of Performance Management on Transport Planning Organizations: Evidence from England. *Transportation Research Record: Journal of the Transportation Research Board*(2119), 105-112.

Marsden, G., & Bonsall, P. (2006). Performance targets in transport policy. *Transport Policy*, 13(3), 191-203.

Marsden, G., Kelly, C., & Nellthorp, J. (2009, April). The likely impacts of target setting and performance rewards in local transport. *Transport Policy*, 16, 59-97.

Marsden, G., Kelly, C., & Snell, C. (2006). Selecting Indicators for Strategic Performance Measurement. *Transportation Research Record: Journal of the Transportation Research Board*(1956), 21-29.

McConnel, K. (2006). Hedonic Wage Analysis. In *Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models (The Economics of Non-Market Goods and Resources)* (pp. 189-242). Springer.

Meyer, M. (1988). Strategic Management in a Crisis-Oriented Environment. *Transportation Research Record*(1156), 88-95.

Meyer, M. (2000). Transport planning for urban areas: A retrospective look and future prospects. *Journal of Advanced Transportation*, 34(1), 143-171.

Meyer, M. (2005). Use of Performance Measures for Surface Transportation in Different Institutional and Cultural Contexts: Cases of Australia, Japan, and New Zealand. *Transportation Research Record: Journal of the Transportation Research Board*, 1924, 163-174.

Miller, T. (2000). Variations between countries in values of statistical life. *Journal of Transport Economics and Policy*, 34(2), 169-188.

Milligan, C., & Kosior, J. (2013). Estimation de la valeur de la vie statistique: examen du cas de la sécurité routière. *23ième Conférence Canadienne Multidisciplinaire sur la Sécurité Routière*. Montréal, QC: L'Association canadienne des professionnels de la sécurité routière.

Milligan, C., Kopp, A., Dahdah, S., & Montufar, J. (2014). Value of a statistical life in road safety: A benefit-transfer function with risk-analysis guidance based on developing country data. *Accident Analysis and Prevention* 71, 236-247.

Milligan, C., Poapst, R., & Montufar, J. (2013). Performance measures and input uncertainty for pedestrian crossing exposure estimates. *Accident Analysis and Prevention*, 50, 490-498.

Miranda-Moreno, L. F., & Fernandes, D. (2012). Pedestrian Injury Risk at Signalized Intersections: Exposure Measures and Geometric Designs. *Transportation Research Board 91st Annual Meeting Compendium of Papers*. Washington DC: Transportation Research Board.

Mizusawa, D., & McNeil, S. (2005). Trinity in transportation planning: Strategic planning, asset management, and performance measures. *Proceedings, Annual Conference - Canadian Society for Civil Engineering*. Toronto, ON: CSCE.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: a nonparametric approach to statistical inference (Sage University Paper Series on Quantitative Applications in the Social Sciences series no. 07-095)*. Newbury Park, CA: Sage.

Moynihan, D. (2005). Goal-Based Learning and the Future of Performance Management. *Public Administration Review*, 65(2), 203-216.

Moynihan, D. (2006). Managing for Results in State Government: Evaluating a Decade of Reform. *Public Administration Review*, 66(1), 77-89.

Mrozek, J., & Taylor, L. (2002). What determines the value of life? A meta analysis. *Journal of Policy Analysis and Management*, 21(2), 253-270.

Navrud, S., & Ready, R. (2007). Review of methods for value transfer. In S. Navrud, & R. Ready (ed.), *Environmental value transfer: Issues and methods*. Dordrecht, The Netherlands: Springer.

NBPD. (2009). *Extrapolation Workbook, Extrapolation Detailed Explanation, and NBPD Facts and FAQs*. Retrieved 11 30, 2011, from National Bicycle and Pedestrian Documentation Project: <http://bikepeddocumentation.org/>

NCHRP. (2005). *NCHRP Report 538: Traffic Data Collection, Analysis, and Forecasting for Mechanistic Pavement Design*. Transportation Research Board.



Nelson, J. P., & Kennedy, P. E. (2009). The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment. *Environmental and Resource Economics*, 345-377.

Nordback, K., Marshall, W. J., & Stolz, E. (2013). Estimating annual average daily bicyclists. *Transportation Research Record*, 2339, 90-97.

Nordback, K., Marshall, W., & Janson, B. (2014). Bicyclist safety performance functions for a U.S. city. *Accident Analysis and Prevention*, 114-122.

Nordfjærn, T., Jørgensen, S., & Rundmo, T. (2012). Cultural and socio-demographic predictors of car accident involvement in Norway, Ghana, Tanzania and Uganda. *Safety Science*, 50(9), 1862-1872.

OECD. (2012). *Mortality Risk Valuation in Environment, Health and Transport Policies*. Paris: OECD Publishing.

Ontario MTC. (1983). *Strategic Management at MTC*. Ontario Ministry of Transportation and Communications. Toronto, ON: Ontario Ministry of Transportation and Communications.

Oxford University Press. (2014, August). *Oxford English Dictionary Online*. Retrieved from OED Online.

Ozbay, K., Bartin, B., Yang, H., Walla, R., & Williams, R. (2010). *Automatic Pedestrian Counter*. Trenton, New Jersey: New Jersey Department of Transportation.

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36(4), 859-866.

Poapst, R., Budowski, A., Vogt, M., & Montufar, J. (2013). *Traffic on Manitoba Highways 2012*. Winnipeg, MB: Manitoba Highway Traffic Information System.

Poapst, R., Hernandez, H., Moshiri, M., MacAngus, J., Reimer, M., Milligan, C., et al. (2012). *Traffic on Manitoba Highways 2011*. Winnipeg, MB: Manitoba Highway Traffic Information System.

Poister, T. (2004). *Strategic Planning and Decision Making in State Departments of Transportation: A Synthesis of Highway Practice*. TRB, Washington, DC.

Poister, T., Margolis, D., & Zimmerman, D. (2004). Strategic Management at the Pennsylvania Department of Transportation. *Transportation Research Record: journal of the Transportation Research Board*, 1895, 56-64.

Pratt, R. H., & Lomax, T. (1996). Performance Measures for Multimodal Transportation Systems. *Transportation Research Record: Journal of the Transportation Research Board*, 1518, 85-93.

Putterill, M., & Maani, K. (1992). Development in performance monitoring systems for roading at the local authority level. *Transportation Research Part A*, 26(6), 471-482.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Australia: R Foundation for Statistical Computing.

Ramani, T., Zietsman, J., Knowles, W., & Quadrifoglio, L. (2011). Sustainability enhancement tool for state departments of transportation using performance measurement. *Journal of Transportation Engineering*, 137(6), 404-415.

Regehr, J., Rådström, B., Arango, J., Isaacs, C., Han, K., Rempel, G., et al. (2006). *Manitoba Highway Traffic Information System: System Maintenance and Upgrade*. Winnipeg, MB: University of Manitoba Transport Information Group.

Reimer, M., & Regehr, J. (2013). Hybrid approach for clustering vehicle classification data to support regional implementation of the mechanistic-empirical pavement design guide. *Transportation Research Record*, 2339, 112-119.

Rosenberger, R., & Loomis, J. (2000). Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation. *Water Resources Research*, 1097-1107.

Rosencrantz, H., Edvarsson, K., & Hansson, S. (2007). Vision Zero - Is it irrational? *Transportation Research Part A: Policy and Practice*, 559-567.

Ruser, J., & Butler, R. (2009). The economics of occupational safety and health. *Foundations and trends in microeconomics*, 5(5), 301-354.

Schneider, R. J., Arnold, L. S., & Ragland, D. R. (2009). Methodology for Counting Pedestrians at Intersections: Use of Automated Counters to Extrapolate Weekly Volumes from Short Manual Counts. *Transportation Research Record: Journal of the Transportation Research Board*, 2140, 1-12.

Schneider, R. J., Henry, T., Mitman, M. F., Stonehill, L., & Koehler, J. (2012). Development and Application of San Francisco Pedestrian Intersection Volume Model. *Transportation Research Board 91st Annual Meeting Compendium of Papers*. Washington, DC: Transportation Research Board.

Schneider, R., Patton, R., Toole, J., & Raborn, C. (2005). *Pedestrian and Bicycle Data Collection in United States Communities: Quantifying Use, Surveying Users, and Documenting Facility Extent*. Chapel Hill, NC: FHWA / Pedestrian and Bicycle Information Center.

Sharma, S. C., Kilburn, P., & Wu, Y. (1996). The precision of average annual daily traffic estimates from seasonal traffic counts: Alberta example. *Can. J. Civ. Eng.*, 23, 302-304.

Short, J., & Kopp, A. (2005). Transport infrastructure: Investment and Planning. Policy and research aspects. *Transport Policy*, 12(4), 360-367.

Stradling, S. G., Anable, J., & Carreno, M. (2007). Performance, importance and user disgruntlement: A six-step method for measuring satisfaction with travel modes. *Transportation Research Part A*, 41(1), 98-106.

Svensson, M. (2009). The value of statistical life in Sweden: Estimates from two studies using the "Certainty Approach" calibration. *Accident Analysis and Prevention*, 430-437.

U.S. FHWA. (2000). *Surface Transportation Efficiency Analysis Model (STEAM 2.0) User Manual*. Retrieved January 28, 2013, from <http://www.fhwa.dot.gov/steam/steammanual20.pdf>

U.S. FHWA. (2009). *Crash modification factors clearinghouse: Install any type of median barrier*. Retrieved January 27, 2013, from <http://www.cmfclearinghouse.org/detail.cfm?facid=42>

U.S. FHWA. (2013). *Crash Modification Factors Clearinghouse*. Retrieved 01 11, 2013, from <http://www.cmfclearinghouse.org/>

University of Manitoba Faculty of Graduate Studies. (2014, August 11). *University of Manitoba*. Retrieved August 11, 2014, from Graduate Thesis/Practicum Guideliens and Info: [http://umanitoba.ca/faculties/graduate\\_studies/thesis/guidelines.html#manuscripts](http://umanitoba.ca/faculties/graduate_studies/thesis/guidelines.html#manuscripts)

US FHWA. (2012, September 11). *MAP-21 - Moving Ahead for Progress in the 21st Century, Performance Management Presentation, September 11, 2012*. Retrieved January 10, 2013, from [http://www.fhwa.dot.gov/map21/docs/11sep\\_perf\\_mgt.pdf](http://www.fhwa.dot.gov/map21/docs/11sep_perf_mgt.pdf)

Vassanadumrondgee, S., & Matsuoka, S. (2005). Risk perceptions and value of a statistical life for air pollution and traffic accidents: Evidence from Bangkok, Thailand. *Journal of Risk and Uncertainty*, 30, 261-287.

Venebles, W., & Ripley, B. (2002). *Modern Applied Statistics with S. Fourth Edition*. New York: Springer.

Viscusi, W., & Aldy, J. (2003). The value of a statistical life: a critical review of market estimates throughout the world. *Journal of Risk and Uncertainty*, 27(1), 5-76.

Viscusi, W., Huber, J., & Bell, J. (2013). Assessing whether there is a cancer premium for the value of a statistical life. *Health Economics*.

WHO. (2012). *Second Global Status Report on Road Safety - Report Flyer*. Retrieved 11 7, 2012, from World Health Organization: [http://www.who.int/violence\\_injury\\_prevention/global\\_status\\_report/flyer\\_en.pdf](http://www.who.int/violence_injury_prevention/global_status_report/flyer_en.pdf)

Willett, J., & Singer, J. (1988). Another Cautionary Note about R2: Its Use in Weighted Least-Squares Regression Analysis. *The American Statistician*, 42(3), 236-238.

Wolfe, A. (1982). The concept of exposure to the risk of a road traffic accident and an overview of exposure data collection methods. *Accident Analysis and Prevention*, 337-340.

World Bank. (2008). *Safe, Clean and Affordable Transport for Development: The World Bank Group's Transport Business Strategy for 2008-2012*. Washington, DC: World Bank Group.

World Bank. (2010a). *Project appraisal document on a proposed loan in the amount of US\$30 million to the Argentine Republic for a road safety project in support of the first phase (APL-1) of a national road safety program*. Washington, DC: World Bank.

World Bank. (2010b). *Project appraisal document on a proposed loan in the amount of US\$250 million to the People's Republic of China for a Ningxia Highway Project*. Washington, DC: The World Bank.

World Bank and WHO. (2004). *World report on road traffic injury prevention*. Geneva: World Health Organization.

World Health Organization. (2014). *WHO Health Statistics and Health Information Systems*. Retrieved 02 21, 2014, from Regional estimates for 2000-2011, Cause Specific Mortality, World Bank Income Groups: [http://www.who.int/entity/healthinfo/global\\_burden\\_disease/GHE\\_DthWBInc\\_2000\\_2011.xls?ua=1](http://www.who.int/entity/healthinfo/global_burden_disease/GHE_DthWBInc_2000_2011.xls?ua=1)

Yang, S., & Davis, G. A. (2002). Bayesian estimation of classified mean daily traffic. *Transportation Research Part A*, 365-382.

Yew, C., & Friedman, P. (2002, Nov-Dec). Measuring the Road to Improvement. *Public Roads*, 66(3).

Yohai, V. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(20), 642-656.

Zall Kusek, J., & Rist, R. C. (2004). *Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington, DC: The International Bank for Reconstruction and Development / The World Bank.

Zegeer, C. V., Stewart, J. R., Huang, H. H., Lagerwey, P. A., Feaganes, J., & Cambell, B. (2005). *Safety Effects of Marked versus Unmarked Crosswalks at Uncontrolled Locations: Final Report and Recommended Guidelines*. McLean, VA: Office of Safety Research and Development, Federal Highway Administration.