Willingness to pay for safety improvements in passenger air travel

Molin, Eric; Blangé, J.; Cats, Oded; Chorus, Caspar

DOI
10.1016/j.jairtraman.2017.04.002

Publication date
2017

Document Version
Accepted author manuscript

Published in
Journal of Air Transport Management

Link to formal publication (Elsevier): https://doi.org/10.1016/j.jairtraman.2017.04.002

Willingness to pay for safety improvements in passenger air travel

Eric Molin (corresponding author)
Transport and Logistics Section
Faculty of Technology, Policy and Management
Delft University of Technology
PO Box 5015, 2600 GA Delft
The Netherlands
Phone: +31 15 278 8510
E-mail: E.J.E.Molin@tudelft.nl

Joey Blangé
MSc-Graduate
Program in Transport, Infrastructure and Logistics
Faculty of Technology, Policy and Management
Delft University of Technology
The Netherlands

Oded Cats
Department of Transport and Planning
Faculty of Civil Engineering and Geosciences
Delft University of Technology
The Netherlands

Caspar Chorus
Transport and Logistics Section
Faculty of Technology, Policy and Management
Delft University of Technology
The Netherlands

Revised Paper submitted for publication to the Journal of Air Transport Management
February 24, 2017

A previous version of this paper was presented at the Annual meeting of the Transportation Research Board, January 8-12, 2017, Washington D.C.

Abstract

The risk of being involved in an airplane accident is largely ignored in air passengers' choice models. The reason presumably is that it is hard to operationalize, because objective safety indicators often involve extremely low probabilities that are hard to grasp and interpret by passengers. In this paper, we propose an operationalization that is based on the perception of safety, which is easy to understand and resonates that perceptions often influence decisions stronger than objective variables. We conceptualize that passengers form a safety perception score of a particular flight based on their perception of airline and route attributes and that this score in turn is traded-off against other flight attributes, such as ticket costs, to arrive at a flight choice. In line with this conceptualization, two stated preference experiments are conducted. In a first experiment, combinations of airline and route attributes are evaluated in terms of safety that is captured on a rating scale. In a second experiment, safety perception is treated as an attribute and traded-off against other flight attributes to arrive at a flight choice. The paper presents the results of a regression and a Panel Mixed Logit model estimated from responses obtained from a convenience sample of 161 air passengers recruited in the Netherlands. The results of both models are then combined to calculate the willingness to pay values for improvements made to a range of airline and route attributes, taking into account socio-demographic variables and psychological traits. As expected, the results indicate that the willingness to pay for improving safety decreases with higher initial safety levels.

Keywords

Safety perception; flight choice; willingness to pay; stated choice experiments; mixed logit model; hierarchical information integration.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

1. Introduction

The risk of being involved in an airplane accident plays an important role in the public discourse surrounding air travel. Such safety risks, however, have been largely ignored in air passengers' flight choice models (Fleischer et al., 2015; Koo et al., 2015, 2016). These models are typically estimated from stated choice experiments that usually only include attributes that can be directly observed, such as ticket price, flight time, access time, number of connections, aircraft-type, and on-time performance (e.g., Hess et al. 2007; Wen and Lai, 2010; Bliemer and Rose, 2011; Brey and Walker, 2011; Collins et al. 2012). To examine to what extent safety is traded-off against those attributes, the inclusion of safety attributes in such an experiment is required. This allows calculating the willingness to pay for improvements in air transport safety, which may be used in managerial decisions and cost-benefit analysis of safety improvements in passenger air transport. Two notable exceptions to ignoring safety related attributes are the studies by Fleisher et al. (2015), who included a safety star rating that essentially represents the risk of being involved in an accident, and by Koo et al. (2015, 2016), who included an attribute describing the number of incidents the aircraft type was involved in in the last three years.

A reason why safety is typically not included in stated choice experiments of flight choice is presumably that it is difficult to operationalize. In safety studies, objective risk indicators are usually formulated in terms of mortality, deaths per number of flights or deaths per distance or unit of time. These indicators, however, often involve extremely rare events that are hard to grasp and interpret by passengers. Moreover, there seems to be a mismatch between such objective risk indicators and how passengers perceive risks (e.g., Kasparson et al., 1988; Savage, 2011). Although air travel is considered one of the safest modes of transport (e.g. IATA, 2016; Squalli and Saad, 2006), Fleisher et al. (2015) found that about a quarter of their respondents suffers moderate to high flying anxiety. This is probably due to passengers' perceived lack of control while traveling by plane and the extensive coverage of airplane accidents by the media who tend to give this more attention than other prominent life threatening events (Rose, 1992; Hall, 2002). However, since it is the risk or safety *perception* of travelers that determine travelers' choices (Boksberger et al., 2007), preferably the operationalization of attributes should be based on passengers' safety perceptions.

That travelers base their choices on their perception of reality instead of on the objectively measurable reality itself, is widely acknowledged in the travel behavior literature (e.g. Koppelman and Pas, 1980; Ben-Akiva et al., 1998; Golledge, 2002; Bonsall et al., 2004). This seem to be even more pertinent with respect to safety, since in contrast to price, travel time, number of transfers etc., information about safety is generally not readily available when booking a flight (Rhoades and Waguespack, 1999; Fleischer et al. 2015). Hence, to form a perception of the safety of a particular flight, passengers have to rely partly on incomplete information and what they believe to be the characteristics of the specific flight. Nevertheless, we assume that they are able to form such safety perception and that this perception can be captured on a rating scale. A safety perception attribute expressed on such a scale can thus be included as an attribute in a choice experiment to examine its trade-off with other flight attributes such as ticket price or flight time, which allows calculating the willingness to pay for improvements in safety perception.

However, such a choice experiment does not reveal any insight into how passengers arrive at their safety perception score, hence, the extent to which various factors influence safety perception. To examine this, we propose to construct a supplementary experiment, a safety perception

experiment. In this experiment, flight profiles that describe a particular flight in terms of attributes that are presumed to influence passengers' safety perception, such as airline and route attributes, are constructed. Respondents are then asked to rate these flight profiles on the same safety perception scale discussed above, which in this experiment is used as the response scale. Hence, the results of this experiment provide insight into which concrete safety measures airlines or flight authorities could employ to improve the perception of safety of passengers. Our proposed approach is inspired by the Hierarchical Information Integration (HII) approach originally proposed by Louviere (1984), which is discussed in the next section.

In sum, this paper contributes to the literature by being the first to simultaneously model (and analyze) the determinants of safety risk perception and the influence of this perception on flight choice behavior. First, the influence of airline and route attributes on perceived safety is examined. Next, the influence of safety perception on flight choice is examined. Unlike previous studies, the proposed modeling approach allows including a potentially large number of safety related attributes and calculate the willingness to pay for improvements made in those attributes, which is an advantage compared with previous modeling approaches that typically included only a single safety related attribute (e.g., Fleischer et al., 2015; Koo et al. 2015, 2016). This approach is demonstrated by reporting the modeling results obtained from a convenience sample of 161 air passengers recruited in the Netherlands.

The remainder of this paper is organized as follows. First, the applied methodology is explained in more detail. Then the modeling results are presented and discussed. First, the results of a regression model estimated from the safety perception rating experiment are discussed. Next, the results of a Panel Mixed Logit model estimated from the stated choice experiment are discussed. For both models, interactions with socio-demographic variables and psychological traits are explored. Then the willingness to pay values for improvements made in route and airline attributes are presented. Finally, conclusions are drawn and relevance for practice is discussed followed by a reflection on the modeling approach.

2. Methodology

As already discussed in the Introduction, we assume that passengers form a safety perception of a particular flight based on their perception of airline and route attributes and that this safety perception in turn is traded-off against objective flight characteristics such as travel costs and time to arrive a flight choice. This assumption implies that a safety related attribute only will affect flight choice if it changes safety perception. This conceptualization is inspired by the Hierarchical Information Integration approach, which is therefore discussed next.

2.1 The Hierarchical Information Integration approach

The Hierarchical Information Integration (HII) approach was originally proposed by Louviere (1984), as a modelling approach for studying complex decisions that involve many attributes (for a review of HII, see Molin and Timmermans, 2009). HII assumes that if faced with a complex decision, decision makers first classify the attributes into a set of higher order decision constructs. Examples of these decision constructs in the transportation context include 'service quality', 'safety' and 'comfort'. In fact, every attribute which is not described in physical terms can be regarded as a decision construct, because an additional experiment is needed to examine how the construct is composed of the

physical terms. It is further assumed that decision makers first form impressions for each of the decision constructs separately, and then integrate these impressions into overall preference values for the decision alternatives or to make a choice.

Consistent with these theoretical assumptions, the implementation of conventional HII models requires the construction of two different experiments. First, a sub-experiment for each construct is required to examine how the attributes defining that construct are traded-off to evaluate the construct. Next, a bridging experiment is required to examine how the decision construct evaluations are traded-off to arrive at an overall evaluation or choice. As an alternative to the rather abstract bridging experiment, Oppewal et al. (1994) prosed to construct integrated HII experiments, which involve constructing a series of experiments where each experiment varies attributes of one decision construct, while describing the other decision constructs in summarizing values. This approach avoids constructing a bridging experiment and offers opportunities for validation.

Yet another variant is applied in Bos et al. (2004) and Molin and van Gelder (2008). In these studies, sub-experiments are constructed to evaluate how physical attributes determine the score of the decision construct. In addition, a decision construct evaluation is included as an attribute in a choice experiment together with physical attributes such as travel costs and travel time. By expressing the levels of the decision construct evaluation on the same scale that is used as the response scale in the sub-experiments, the results of the model estimated from the sub-experiment and the choice experiment can be linked. This modeling approach is also applied in this paper, though in this paper we assume only a single decision construct, namely, safety perception.

Our modeling approach assumes that passengers consider airline and route characteristics to arrive at their perception of safety of a particular flight, which is then traded-off against other more readily available objective flight characteristics such as ticket price and travel time. In line with this assumption, we construct two different experiments to incorporate safety perception in flight choice models. The first experiment examines how safety perception is affected by the airline and route characteristics. The second experiment examines how perceived safety is traded-off against observable travel attributes such as ticket price and flight time. The following two subsections discuss these two experiments in more detail.

2.2 The flight safety perception experiment

The flight safety perception experiment intends to measure how passengers perceive the safety of a particular flight. Safety perception may be influenced by a large number of factors, which may involve attributes that passengers associate with airplane accidents based on own experience, stories from persons of their social network or influenced by media coverage of airplane accidents. This potentially encompasses airline attributes, such as carrier type, previous involvement in accidents, age of fleet, home country, and route and other flight conditions, such as bad weather conditions and flying over water, mountain areas or conflict areas. Hence, in the safety rating experiment, combinations of such attributes (profiles) are presented to respondent and they are requested to evaluate the safety of these profiles.

In order to avoid overloading respondents with too many attributes, we limited the number of attributes to be included in the safety perception experiment. From a larger list of potential attributes, we selected six attributes that in our opinion overall scored best on the following three criteria: (i) passengers should have access to information about the factor; (ii) the attributes are actionable, that is they can be controlled or changed by airlines; (iii) the value of the attributes can

differ between flights, which excluded common causes of airplane accidents such as bird strikes (Thorpe, 2003), runways and airports.

The following six attributes were selected:

1. Airline Safety Index: Following Chang and Yeh (2004) we included a hypothetical airline safety index, that expresses an overall safety score of an airline based on the airline's performance on management, maintenance, operations and planning with regard to safety. Respondents have to assume that the regulatory organization of commercial airlines (IATA) would issue such an index. Similar to the car industry and similar to Fleischer et al. (2015), the safety index is expressed in stars. We arbitrarily selected 1 to 4 stars as attribute levels. Respondents are explained that all airlines that have a star rating are allowed to execute flights and are not blacklisted, but they differ in levels of safety: one star means that the airline complies with safety regulations but performs at the minimum required level, while four stars means that the airline conducts a range of extra measures and maintenance to maximize safety. Including this attribute allows testing how such an objective safety index determined by an independent air authority would affect safety perception if it would be implemented.
2. Carrier type: Many passengers are aware that low-cost carriers save on all services that do not have immediate commercial value, which they may project on safety procedures. More specifically, they may be aware that low-cost carriers focus on fast turn-around times and some may be aware of their reputation to take as little fuel with them as possible to save money, which nearly led to accidents as airplanes almost completely ran out of fuel (Zhou and Drukier, 2015). Hence, this attribute examines whether passengers perceive low-cost carriers and full-service carriers differently in terms of safety.
3. Number of accidents with fatalities (Kebabjian, 2015): It is assumed that an airline being involved in accidents affects the safety perception (attribute levels: 0, 1, 2, 3 accidents in the past 10 years). This attribute has some similarity to the safety attribute operationalized by Koo et al. (2015, 2016) as the number of accidents (0, 1 or 3) the aircraft type was involved in, in the last three years.
4. Flying over water: There are quite some crashes that have happened above water, which may influence the fear of loss of external control among passengers (Foreman and Van Gerwen, 2008).
5. Flying over conflict-areas: This factor became of interest after the accident with MH17 flying from the Netherlands to Malaysia that crashed in the conflict area of Ukraine in the year preceding our data collection.
6. Bad weather conditions: Although this factor does not meet the criteria for selection discussed before (it is typically unknown when booking a flight, it does not differ between different flights operated at the same time, and it cannot be controlled by airlines), we made an exception for this factor, as we expect that many passengers perceive this to be a major cause of airplane accidents.

A main-effects only fractional factorial design was selected to construct the profiles that describe combinations of these attributes. This resulted in 16 profiles, which were blocked in two sets of 8 attribute level balanced profiles. Hence, each respondent was randomly assigned to only one block of 8 profiles. Respondents were requested to evaluate each of those profiles in terms of safety and express their evaluation on a seven-point rating scale ranging from (1) very unsafe to (7) very safe. Figure 1 presents the example profile and explanation of the safety perception rating task as included in the questionnaire.

In this part of the survey, you have to imagine you are making an intercontinental flight to a far holiday destination. We present 8 different flights to you, which vary in some characteristics of the airline and the route to be followed. We ask you to judge every situation based on how safe you think the flight is. Express your judgment on a scale from (1) very unsafe to (7) very safe.

Example flight

Safety score of the airline: 3 stars (out of 4)

Airline type: full-service carrier

Number of accidents with fatalities of this airline in the past 10 years: 2

Will fly over water

Will fly over conflict areas

Stormy weather circumstances

		1	2	3	4	5	6	7	
Example flight	(Very unsafe)	○	○	○	○	○	○	○	(Very safe)

Figure 1. Flight safety perception rating task example

2.3 The flight choice experiment

The aim of the flight choice experiment is to examine how safety perception is traded-off against other flight attributes such as ticket costs. Because the main goal of our study is to calculate the willingness to pay for safety improvements, we limited the number of attributes to describe the flight choice alternatives to the following four attributes (the selected levels added in parenthesis) in order to keep the number of choice sets to be evaluated by respondents at a minimum:

(i) travel costs, which is varied in the levels 600, 800 and 1000 euro;

(ii) travel time, which is varied in the levels: 11, 13 and 15 hours;

(iii) comfort, which is varied in the levels:

- Low (Little legroom, little service (additional payment for certain meals/drinks))
- Medium (Much legroom, but little service (additional payment for certain meals/drinks))
- High: (Much legroom and full service on board (all inclusive, no additional payments)).

(iv) safety perception, which is varied in the levels: 1 = very safe; 4, and 7= very unsafe.

The attribute *safety perception* expresses the respondent's perception of safety of that particular flight. Hence, respondents have to assume that the presented level is their own evaluation of the airline and route attributes of that particular flight. This attribute was explained to the respondents in the questionnaire as follows: "This rating is the safety judgment as if you would judge the associated flight. This matches with the ratings you provided in the previous part of this survey, based on the varied characteristics of the airline and route. The rating is expressed on the same 7-point rating scale, where 1 stands for 'very unsafe' and 7 stands for 'very safe'." By referring back to the safety perception rating experiment, this implies that that experiment needs to be completed by the respondents before they conduct the choice experiment. It is important that the safety attribute is expressed on the same scale that is used as the response scale in the safety perception rating experiment, because this allows linking the results of the models estimated from either experiment. Similar to the other three attributes, the number of levels selected to vary safety perception is

limited to three, for which we selected both extreme scale values, hence the levels 1 and 7 and the middle value, 4. Effects of the other values can be found by linear interpolation once the model is estimated.

To assure that all respondents had the same type of trips in mind while making flight choices, the context they had to assume was defined as “an intercontinental flight to a faraway holiday destination”. This offers a context for which respondents can imagine that all levels of the airline and route attributes that were varied in the safety perception rating experiment can occur. Moreover, limiting the context to a recreational trip instead of a business trip, has the advantage that it may be assumed that respondents pay themselves for the flight tickets and not the company they work for, which is likely to increase the validity of the willingness to pay results.

The selected attributes and levels resulted from fine tuning the experiment based on the results of two pilot surveys involving 89 and 39 respondents respectively. For the second pilot study, choice sets of two unlabeled alternatives each were constructed from an orthogonal fractional factorial design involving 18 choice sets. From the observed choices in that study a Multi Nomial Logit (MNL) model was estimated of which the estimated parameters were used as priors for constructing a D-efficient design for the experiment of the main survey. In total, 18 choice sets were constructed, which were blocked into two sets of 9 choice sets each, each having attribute level balance. Every respondent was randomly assigned to complete one block of 9 choice sets. The respondents made these choices after they rated 8 profiles of the safety perception rating experiment. Figure 2 presents the choice set example and explanation that was included in the questionnaire.

Part 2: Flight choice experiment

In this second part of the survey, you will see 9 choice problems between 2 flights which both go to a far intercontinental holiday destination. Mark the flight of your choice by ticking the circle below the preferred flight.

Characteristics	Example flight A	Example flight B
Price	€ 1000	€ 600
Comfort	medium	low
Your safety rating	7 out of 7	1 out of 7
Travel time	15 hours	11 hours
	◯	◯

Figure 2. Flight choice task example

2.4 Integration of the two experiments

Figure 3 graphically displays the structure of the two experiments and the two functions that can be estimated from those experiments. The safety perception rating experiment examines how safety perception is influenced by the six airline and route attributes, which is captured in mapping function f. The parameters of the selected six attributes can be estimated from the observed safety perception scores in this experiment and indicate the relative weight each of the six attributes has in determining the perceived safety score of a particular flight.

The flight choice experiment examines how safety perception is trade-off against other flight attributes, which is captured in mapping function g. The parameters estimated from the observed choices between flight alternatives indicate the relative weight the four selected flight attributes (travel time, travel cost, perceived safety and comfort) have in determining the utility U derived from a flight alternative. Note that in the safety perception experiment, perceived safety is an observed variable, while in the choice experiment, safety perception is an attribute, the values of which are determined by a statistical design.

After the functions f and g are estimated from the conducted experiment, they are both required to predict the utility of a particular flight alternative. First, the perceived safety value of a particular flight is based on function f. This predicted value is than treated as the attribute value of the perceived safety attribute in function g. The dashed line In Figure 3 symbolizes this link.

Later on, we show how the results of the two experiments can be combined to calculate the willingness to pay values for improvements made in the attributes that influence the perception of safety.

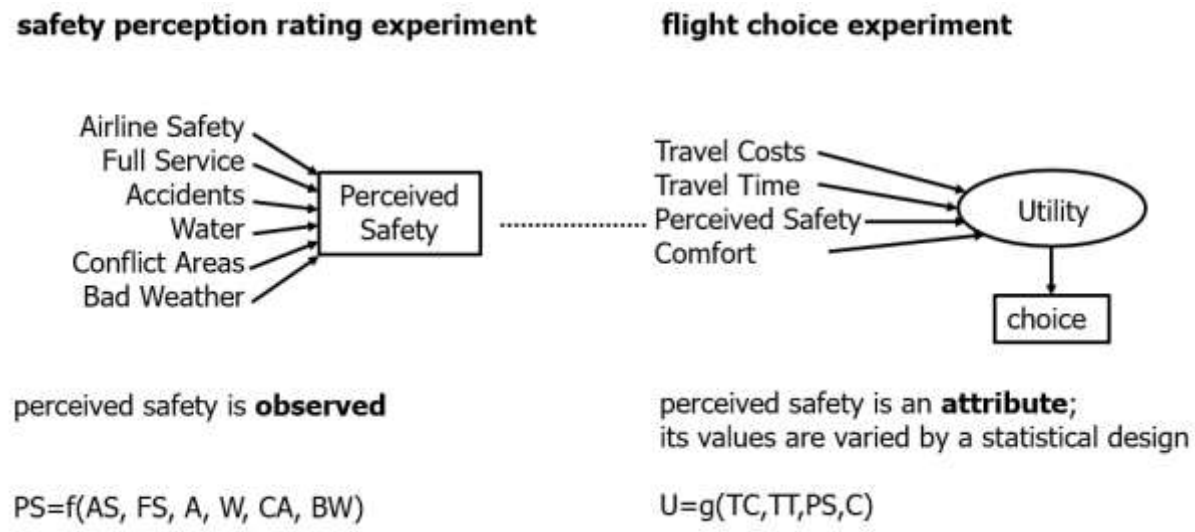

Figure 3. A graphical display of the two experiments

2.5 Background variables

In the third part of the survey, a range of personal characteristics was measured. This involved (i) socio-demographic variables: gender, age, income, business traveler, living abroad; (ii) air travel related characteristics: number of intercontinental flights made in past five years and involvement in (near) aircraft accidents; (iii) psychological traits: *fear of flying* and risk-taking attitude. *Fear of flying* is measured by the single item statement "I have fear of flying", for which a five-point Likert scale is used as a response variable (1=strongly disagree to 5= strongly agree). Risk-taking attitude is measured by the average score of the likelihood of being engaged in 8 different risky recreational activities, which was taken from the domain specific risk attitude test devised by Weber et al. (2002). An example of such a risky activity statement is "going camping in the wilderness, beyond the civilization of a campground" (1= highly unlikely, 5=highly likely).

2.6 Data collection and sample

The population was defined as all individuals aged 18 years or older who have made at least one single intercontinental flight within the past five years. A snowball sampling method has been applied to recruit respondents starting from the second author's social network. He asked ten persons from his personal network from different age groups and work environments to fill out the questionnaire and additionally asked them to send an invitation to fill in the online survey to all persons from their address book. This resulted in a sample of in total 161 respondents who completely filled out the questionnaire in the Fall of 2015.

Table 1 Distribution of background characteristics (N=161)

Gender	
Male	59.0%
Female	41.0%
Age	
18-22	14.9%
23-32	26.7%
33-42	14.3%
43-52	15.5%
53-62	19.9%
63-72	8.7%
Net yearly income in 100,000 euros	
0-0.2	32.9%
0.2-0.4	26.7%
0.4-0.6	20.5%
0.6-0.8	6.2%
O.8-1.0	2.5%
1+	4.1%
Does not know or is not willing to tell	9.1%
Business traveler	
also travels for business purposes	19.9%
does not travel for business purposes	80.1%
Country of residence	
Netherlands	90.1%
Outside Netherlands (abroad)	9.3%
Number of Intercontinental leisure flights in the past 5 years	
0	0%
1-5	85.1%
6-10	9.9%
11-15	3.7%
16-20	1.2%

Table 1 presents the frequency distribution of the background variables. The results show that a relatively even distribution is realized among the distinguished categories. Although there is a slight overrepresentation of males, there does not seem to be a disturbing over- or underrepresentation of other specific groups. However, this sample cannot be considered a

representative sample given the non-random recruitment process, hence, this should be considered a convenience sample. Nevertheless, we believe that it is suitable for a first exploration of the impact of safety perception on flight choices. Nonetheless, care should be taken in generalizing the results to the larger population of air passengers.

2.7 Model estimation procedure

In this section, we discuss the model estimation procedures applied for the rating and the choice experiments. The dependent variable in the model estimated from the rating experiment is the perceived safety rating observed for each profile. This rating was measured on a 7-point scale, which was discussed earlier. One may argue that this is an ordinal scale, which would imply that an ordered logit model should be estimated. Being consistent with this assumption would imply that the safety perception attribute that was varied in the choice experiment should also be treated as an ordinal variable in the choice model estimated from that experiment. Consequently, this attribute should enter the choice model as a series of dummy variables which would not allow interpolation for unselected intermediate values. As we wish to achieve the latter, we need to assume that the 7-point safety perception scale is of interval measurement level, which allows estimating linear parameters for the safety attribute. Also for reasons of parsimony, this is the preferred option. Consistently, we assume that the rating scale of the rating experiment is of interval level measurement. This allows us to estimate an ordinary linear regression model from the observations made in the safety perception rating experiment. In addition to the parameters estimated for the varied attributes, we explored to what extent these parameters differ between the categories of background variables. Therefore, we tested which of the interactions of the background variables and attributes were statistically significant. This resulted in the following estimated function to predict the perceived safety score for an airline and route profile *j*:

$$PS_j = C + (\beta_{AS} + \beta_{AS}^{Age} \cdot Age) \cdot AS_j + (\beta_A + \beta_A^{RA} \cdot RA) \cdot A_j + (\beta_{FS} + \beta_{FS}^{IA} \cdot IA) \cdot FS_j + \beta_{BW} \cdot BW_j + (\beta_{CA} + \beta_{CA}^{Age} \cdot Age) \cdot CA_j + \beta_W \cdot W_j + \beta_{Age} \cdot Age + \beta_{IA} \cdot IA + \beta_{RA} \cdot RA + \beta_{RTA} \cdot RTA + \beta_{Male} \cdot Male + \beta_{NoF} \cdot NoF + \beta_{FoF} \cdot FoF$$

Where: PS=predicted perceived safety score, C=regression constant, AS=Airline Safety Index, A=Number of Accidents the airline was involved in, FS=Full Service Carrier dummy, BW=flying over Bad Weather dummy, CA=flying over Conflict Area dummy, W=flying over water dummy, RA=respondent resides abroad dummy, IA=Involvement of respondent in Accidents, NoF=number of flights the respondent made in the past 5 years, FoF=Fear of Flying. The estimated coefficients and their t-values are presented in Table 2.

From the choices observed in the choice experiment, we estimated a series of logit models to explore interactions with background variables. We found significant interactions for income and risk-taking attitude with the price parameter. This resulted in the following specification of the utility function: the utility of alternative i for an individual from income class k and risk-taking class m, is specified as follows:

$$U_i = V_i + \varepsilon_i =$$

$$-\beta_{TC} \cdot TC_i + \beta_{TC}^{Inc} \cdot Inc_k \cdot TC_i + \beta_{TC}^{Risk} \cdot Risk_m \cdot TC_i + \beta_{PS} \cdot \ln(PS_i) + \beta_{TT} \cdot TT_i + \beta_C \cdot C_i + \varepsilon_i$$

Note that perceived safety (PS) enters utility in a logarithmic form, reflecting a decreasing marginal utility of perceived safety (i.e., the higher the initial level of perceived safety, the smaller is the

positive impact on utility of an additional increase in perceived safety). The logarithmic specification had a substantially and significantly better fit than a linear specification (see Table 3), and an equally good fit as a linear+quadratic term specification (which required an additional parameter to be estimated). In an attempt to adequately capture random taste heterogeneity and panel effects (recall that each respondent made multiple choices) we tested a range of distributions for each taste parameter. The optimal combination of distributions for random parameters was found to be as follows: $\beta_{TC} \sim LN(\mu_{TC}, \sigma_{TC})$, $\beta_{PS} \sim U(B_{PS} - S_{PS}, B_{PS} + S_{PS})$, $\beta_C \sim U(B_C - S_C, B_C + S_C)$. The choice for the LogNormal distribution for β_{TC} was made to ensure a negative sign for $-\beta_{TC}$ in the utility function, implying a dislike of higher costs. Note that the estimated μ_{TC}, σ_{TC} refer to the corresponding Normal distribution. The mean and standard deviation of the LogNormal distribution itself are obtained as follows: mean=$\exp(\mu_{TC} + (\sigma_{TC})^2/2)$ and st. dev.=$\sqrt{(\exp(\sigma_{TC})^2 - 1) \cdot (\exp(2 \cdot \mu_{TC} + (\sigma_{TC})^2))}$. In addition to the LogNormal distribution, a uniform distribution with positive support was tried as well, with unsatisfactory results (slow and unstable convergence). For β_{PS} and β_C a Uniform distribution provided the best results in terms of model fit and stability of convergence; the midpoint B and half-range S of the distribution were estimated, so that the distribution itself is characterized as $U(B - S, B + S)$. Also for these distributions, an attempt was made to constrain them to the positive domain (by imposing the constraint that $B = S$), with no good results in terms of model fit and stability of convergence; As a consequence, these two estimated random parameters end up with a small probability mass in the negative domain (5% for β_{PS}, 6% for β_C). We accept these small counterintuitive probability masses in exchange for being able to accommodate the substantial levels of unobserved heterogeneity associated with taste for these attributes in our sample. We did not find significant levels of unobserved heterogeneity for the travel time parameter, and hence have decided to estimate it as a crisp parameter. We also decided not to allow for random sociodemographic and attitudinal effects (β_{TC}^{Inc} and β_{TC}^{Risk}) for reasons of parsimonity and model stability. We used 4000 Halton draws to estimate the Mixed Logit (ML) model. Python Biogeme (Bierlaire, 2016) was used for estimation. Model estimation results are reported in Table 3;

Willingness to pay (WtP) is defined in terms of the ratio of partial derivatives of systematic utility with respect to the considered attribute and the cost attribute (TC). For an individual from income class k and risk-taking class m, this leads to the following formulations of WtP for Perceived safety (PS), Travel time (TT), and Comfort (C):

$$WtP_{PS} = \frac{\frac{\partial V}{\partial PS}}{\frac{\partial V}{\partial TC}} = \frac{\beta_{PS}/PS}{\beta_{TC}^{Inc} \cdot Inc_k + \beta_{TC}^{Risk} \cdot Risk_m + \beta_{TC}}$$

$$WtP_{TT} = \frac{\frac{\partial V}{\partial TT}}{\frac{\partial V}{\partial TC}} = \frac{\beta_{TT}}{\beta_{TC}^{Inc} \cdot Inc_k + \beta_{TC}^{Risk} \cdot Risk_m + \beta_{TC}}$$

$$WtP_{C} = \frac{\frac{\partial V}{\partial C}}{\frac{\partial V}{\partial TC}} = \frac{\beta_{C}}{\beta_{TC}^{Inc} \cdot Inc_k + \beta_{TC}^{Risk} \cdot Risk_m + \beta_{TC}}$$

Note that since PS entered the utility function in logarithmic form, its WtP depends on the initial level of PS as can be seen in the equation for WtP_{PS}; higher PS leads to a lower marginal utility of a further increase in PS, and as such to a lower WtP. Also remember that $\beta_{PS} \sim Uniform$ and $\beta_{TC} \sim LogNormal$; this implies that we must obtain (mean) estimates of WtP by means of simulation. This is done as follows: for each parameter 25,000 draws are made from the corresponding distribution (experimentation with various numbers of draws suggests that the used number is sufficient); for each pair of draws (i.e., the perceived safety attribute and the price attribute) the associated WtP is computed. The average of these WtP-draws then constitutes the mean WtP for the attribute. We also report the median, and the 10th and 90th percentile WtPs. For PS, we report WtP for each initial level of PS except for the highest level of seven (for which no further improvement is possible). Results are presented in Table 4, where we focus on the median income class (0.3 representing the income class 20.000 to 30.000 euro) and the median level of risk-taking (2.25); note that for this combination (like for most other combinations) of income and risk-taking, a positive sensitivity for travel costs (i.e., a dislike for costs) is guaranteed, such that the ratios in the above equations are defined for the full range of $\frac{\partial V}{\partial TC}$.

3. Results

In this section, we present and discuss the modeling results. First, we present the result of the safety perception experiment and discuss how airline and route attributes influence safety perception. Next, we present the results of the flight choice experiment and discuss how safety perception is traded-off against other flight choice attributes. Then the results of both experiments are combined to calculate the willingness to pay for improvements in airline and route characteristics. Finally, we compare our results with those of previous studies.

3.1 The safety perception model

Table 2 presents the results of the regression model estimated from the safety perception rating experiment. The dependent variable is the observed safety perception rating and the independent variables are the six airline and route attributes that are varied in the rating experiment. In addition, several interaction effects of parameter with socio-demographic variables and psychological traits were found to be statistically significant.

The estimated regression coefficients indicate to what extent the safety perception rating changes with one unit change of the predictor variable. The results of Table 2 can be interpreted as follows:

- Airline safety index (AS) (0.526) seems to have a relatively large impact on safety perception, however, its effect needs to be interpreted together with its interaction with age (-0.004). The latter suggests that the impact of an objective safety index decreases with age. To illustrate: the AS parameter of an 18 year-old is 0.526 + 18*-0.004= 0.454, whereas the same parameter for a 70 year-old is only 0.246. Hence, a flight by a four star airline (highest safety standard) compared to a flight by one star airline increases the safety perception (ceteris paribus) of a 18 year old passenger by 1.362 points, while for a 70 year old passenger the increase is only 0.738 points.
- Also the number of accidents with fatalities an airline is involved in in the past ten years has a relatively large impact (-0.303): the difference in safety perception score between zero accidents

and three accidents results in a decrease of 0.909 safety perception points. The interaction effect with living abroad (-0.301) suggests that this impact is twice as large for those living abroad.

- The parameter for flying over conflict areas seems to have a counterintuitive sign (0.404), but when its interaction with age (-0.025) is taken into account, it becomes clear that the impact is negative for all age groups. To illustrate: the safety score impact of flying over conflict areas for a 18 year old is 0.404 + 18 * -0.025= -0.046, which may be considered a rather limited impact; whereas for a 70 year the effect is -1.346 points, which is substantial.
- The three remaining attributes all have expected signs. The safety perception of a flight decreases with bad weather conditions (-0.540) and flying over water (-0.146), and increases if the airline is a full service carrier (0.264). The interaction of the latter attribute with the personal characteristic *being involved in (near) accidents* (0.787) suggests that for this group a flight with a full service carrier substantially increases the safety perception of a flight.
- The results for the background variables indicate to what extent the average safety perception of flights varies among background variable categories. The overall safety perception increases with age, living abroad, risk-taking attitude, being male, and with increased frequency of flying, whereas it decreases with being involved in (near) accidents and with fear of flying.
- Except for the results for living abroad, for which it is not directly clear what the mechanism is that causes this effect, all reported results seem to be plausible.

Table 2 Estimates of the safety perception model (linear regression)

	Estimate	*t*-Ratio
regression constant	2.837	8.602
Airline Safety (1-4)	0.526	5.996
Airline Safety * Age	-0.004	-1.969
Number of Accidents (0-3)	-0.303	-8.806
Number of Accidents * Residence Abroad	-0.301	-2.681
Full Service (dummy)	0.264	3.542
Full Service * Involved in Accidents	0.787	2.031
Bad Weather (dummy)	-0.540	-7.365
Conflict Areas (dummy)	0.404	2.072
Conflict Areas * Age	-0.025	-5.509
Water (dummy)	-0.146	-1.995
Age (18-70)	0.043	4.912
Involved in Accidents (dummy)	-1.454	-2.369
Residence Abroad (dummy)	0.813	2.623
Risk-Taking Attitude (1-5)	0.224	4.523
Male (dummy)	0.304	3.812
Number of Flights (1-50)	0.020	3.576
Fear of Flying (1-5)	-0.076	-2.089

R^2*=0.296*

3.2 The flight choice model

Table 3 presents the estimates of two MNL models (one with a linear specification of PS, and one with a logarithmic specification) and the Panel Mixed Logit model estimated from the choices observed in the flight choice experiment. The results can be summarized as follows:

- All attributes have significant parameters and the directions of the signs are as expected: higher Comfort and higher Perceived Safety levels increase utility, while higher Travel Costs and higher Travel Time decrease utility.
- Comparing MNL-1 with MNL-2 indicates that a parameter estimated for the natural logarithm of Perceived Safety improves the model fit (more than 10 points difference with the same number of parameters) in comparison to linear parameters. This means that the utility increase per unit diminishes with higher initial values of perceived safety. This conform our intuition: the higher the Perceived Safety level already is, the less a further increase contributes to utility.
- As earlier reported, the estimated sigma's for Comfort, Perceived Safety and Travel Cost are statistically significant, suggesting heterogeneity in the valuation of Perceived Safety, Travel Costs and Comfort, while the sigma for Travel Time is not statistically significant, which suggests homogeneity in Travel Time valuation. Note that sigma (Travel Costs) represents the standard deviation of the underlying Normal distribution. Sigma (Perceived Safety) and sigma (Comfort) represent the half-range of the Uniform distribution; see the Section "Model estimation procedure" for details.
- As is expected given the three significant random parameters, the Panel Mixed Logit model achieves a much higher model fit than the corresponding MNL model (the difference is highly significant, as indicated by a Likelihood Ratio Statistic which is more than seven times larger than the critical Chi-squared value at a 1% significance level (3 degrees of freedom). This suggests that capturing taste heterogeneity and correlation between choices made by the same individual has a profound positive impact on model performance.
- The positive value of the interaction between Travel Costs and Income indicates that costs weighs less with increasing income, which is the expected direction. The negative value of the interaction between Travel Costs and Risk-Taking attitude suggests that costs are more important with higher risk-taking attitude. This result suggests, as can be expected, that risk-taking persons are less willing to pay for arriving at higher safety levels.
- Based on the estimates of the linear cost parameter as estimated by MNL-1, its value for the median income (0.3) and median risk-taking class (2.25) can be calculated as -0.396. Given the reported Travel Time parameter of this model, the WtP value for saving 1 hour of travel time can be calculated as 58.33 euro. This is similar to the value of 51.75 euro per hour saved travel time for air travel found by the official Dutch value of time study (KIM, 2013). Hence, this finding gives confidence in the results of our experiment.

Table 3 Estimates of the flight choice model

	MNL-1		MNL-2		Panel Mixed Logit	
	Est.	*t*-Ratio	Est.	*t*-Ratio	Est.	*t*-Ratio
Comfort (β_C)	0.563	10.11	0.502	8.63	0.834	9.11
Perceived Safety (β_{PS})	0.475	17.38				
ln(Perceived Safety) ($\beta_{ln(PS)}$)			1.49	17.27	2.66	11.22
Travel Time (β_{TT})	-0.231	-9.21	-0.244	-9.69	-0.371	-10.68
Travel Costs (β_{TC})	-0.241	-2.88	-0.244	-3.02	-0.993	-2.75
Travel Costs * Income (β_{TC}^{INC})	0.280	2.88	0.270	2.87	0.458	2.93
Travel Costs * Risk-taking (β_{TC}^{RISK})	-0.106	-3.72	-0.103	-3.72	-0.126	-2.69
sigma Comfort (σ_C)					-0.951	-3.92
sigma Perceived Safety ($\sigma_{ln(PS)}$)					2.93	9.54
sigma Travel Costs (σ_{TC})					0.710	3.44
Parameters		6		6		9
Final Loglikelihood		-668.068		-657.576		-602.094
Rho-square		0.335		0.345		0.401

scaling: TC = observed ticket price / €100; INC = observed income category / €100.000

3.3 Willingness to pay for safety improvements

Table 4 presents the WtP values for one point increase on the safety perception rating scale for the median income and risk-taking class given an initial safety perception value, which are obtained by simulation as earlier discussed. Included are the values based on the MNL-2 model and those based on the ML model obtained by simulation as earlier discussed for which we present the median value, the average and the 10th and 90th percentiles. As is usually the case, WtP estimates from the MNL model are somewhat lower than the median and mean values obtained by the Panel Mixed Logit model. In light of the fact that the Mixed Logit model had a much better fit with the data, we prefer its WtP estimates over those of the MNL model. Like previous studies (e.g. Brownstone and Small, 2005), we use the Mixed Logit model's median estimates for subsequent analyses. Table 4 illustrates that WtP values decrease with higher initial safety perception levels, which is completely in line with estimated parameters and intuition.

Table 4 Willingness to pay values (in euro's) for the median income and risk-taking class

Initial_PS	WtP MNL-2	Median_WtP ML	Mean_WtP ML	10th percentile	90th percentile
1	385	448	552	51	1196
2	192	224	276	26	598
3	128	149	184	17	399
4	96	112	138	13	299
5	77	90	110	10	239
6	64	75	92	9	199

Table 5 Median willingness to pay values (in Euros) for safety improvements (for the median income and median risk-taking class)

	initial safety perception level					
	1	**2**	**3**	**4**	**5**	**6**
one safety index score increase						
age=18	203	102	68	51	41	34
age=36	171	86	57	43	34	29
age=54	139	69	46	35	28	23
age=72	107	53	35	27	21	18
Full service carrier						
never involved in accident	118	59	39	30	24	20
involved in accident	471	235	157	118	95	79
airline with 1 accident less						
lives in Netherlands	136	68	45	34	27	23
lives abroad	271	135	90	68	54	45
avoid flying over water	65	33	22	16	13	11
avoid flying over conflict areas						
age=18	21	10	7	5	4	3
age=36	222	111	74	56	45	37
age=54	424	212	141	106	85	71
age=72	625	313	208	156	126	105
avoid flying in bad weather	242	121	80	60	49	41

Now the results of both models can be combined. First, the estimated parameter for an airline or route attribute of the safety perception model as presented in Table 2 , indicates how many points the perceived safety score changes by one unit increase of that attribute. Next, the Median WtP ML values presented in Table 4 indicate how much euro a median passenger is willing to pay to increase one point in safety perception level, given a certain initial safety level. Hence, by multiplying an airline or route attribute parameter with these WtP values, the WtP values for an improvement in those attributes can be obtained. Thus, the WtP to improve one unit of change in airline or route attribute *a* for the median income and risk-taking class can be calculated as follows (taking the absolute of the coefficient assures that the sign of the WtP value is correct):

$$WtP_a = \left|\beta_a\right| \cdot median\ WtP_{PS}\ ML$$

We provide an example by calculating the WtP-values for avoiding to fly over a conflict area of a 36 year-old individual from the median income and median risk-taking class, and whose initial safety perception level is 3. Table 4 indicates the corresponding WtP-value is 149 euro. Table 2 indicates that avoiding to fly over a conflict area increases the safety perception score by $|(0.404 \cdot -0.025 \cdot \text{age})|$, which for a 36 year old reveals the value 0.496. Hence, the WtP value to avoid flying over conflict areas is equal to 0.496 · 149=74 euro, which is the number presented in Table 5. This Table provides the calculated WtP values made for one unit change in the six airline and route attributes, taking into account the initial safety perception level and background variables of the passenger. Completely in line with earlier discussed results, Table 5 illustrates that the WtP values decrease with

higher initial perceived safety values. Note that this table only presents the results for the median WtP values that apply for the median income and risk-taking class.

3.4 Discussion of the results

In this section, we compare the results of our study with results of previous research. One of the attributes we included in the safety rating experiment, 'the number of accidents with fatalities the airline was involved in the last 10 years', was fairly similar to the attribute Koo et al. (2015, 2016) included as the only attribute indicating safety information in their stated choice experiment. This involved the number of incidents (0, 1, or 3) the aircraft type had in the last three years. Over the complete range of incidents, it can be calculated from the presented results in their study that passengers are willing to pay 224 euro to fly with an airplane type which has one accident less. This is less than the value of 271 euro we find in our study for those living abroad but more than the 136 euro the majority of the respondents is willing to pay for one accident less (see Table 5), assuming low initial safety levels. The somewhat larger values found in the Koo et al. (2015, 2016) study may be explained by the fact that the incidents in their study refer to the last three years, while ours refer to a longer time period of 10 years. Nevertheless, that the results are somewhat in the same range gives confidence in our calculated WtP values.

Our attribute 'Airline Safety Index' is fairly comparable to the only safety information attribute Fleisher et al. (2015) included in their study. They also used an objective safety value determined by a flight authority and expressed their levels in safety stars. They found in their research no significant difference in utility between 1 and 2 stars (low and medium safety levels) and only a significant difference between either of these two levels and three stars (high safety level). The results of our research, however, tell a different story. The average observed safety ratings of the airline and route profiles grouped on safety index levels were the following: 3.70 for one star, 4.16 for two stars, 4.33 for three stars, and 4.82 for four stars. A post-hoc ANOVA test revealed significant differences between all pairs of levels, except between the second and the third level. Furthermore, we find a significant linear relationship between the safety index attribute and perceived safety, while the coefficient estimated for the quadratic component is negligible and far from reaching statistical significance. Our results suggest that though passengers may have troubles distinguishing between the two middle values, they are well able to distinguish those two values from both the highest and the lowest safety rating index values. Possibly, the limited number of levels used by Fleisher et al. (2015) caused their results. Another possibility is that in Fleisher et al. (2015) the objective safety score varied freely within each airline company (El Al, Lufthansa, EasyJet and Earoflot). Hence, a particular company may be paired with one star in one alternative and with three stars in another alternative, which may even appear in the same choice set. It was explained to respondents that this was caused by the flights making use of different airplane types. However, in our opinion this does not match with the earlier explanation provided to the respondents, that involved that the safety rating was determined by a flight authority and was based on "…the history of accidents, near-accidents, and pilot errors, the quality of aircrew training and expenditures on aircraft maintenance." (Fleisher et al., 2015, p. 213), which implies that this a relatively stable score. Thus, a safety score is airline specific and should therefore not vary freely between different flights conducted by the same company flying under the same brand name. This has possibly affected the results. In the light of this discussion, the conclusions Fleisher et al., 2015, draw in their study: "…people are not sensitive to the different shades of safety, and instead, they simply discern flights as either safe or unsafe...", seems premature and too far reaching.

Furthermore, our finding that full service carriers are perceived as safer than low-cost carriers, is consistent with the perception ratings of two Western-European airlines reported by Fleisher et al. (2015). They find that a full-service carrier (Lufthansa) scored on average 5.4 on a 7 point rating scale, while a low-cost carrier (EasyJet) scored 4.2 on average on the same scale.

Put in a broader perspective, this study addresses some of the issues raised in the literature. Koo et al. (2016) wondered what people mean by the term safety. Our results shed a bit of light on how this construct is conceived by showing which attributes influence safety perception. The results suggests that safety is not only directly influenced by presented objective risk probabilities of being involved in accidents, but also by a range of other airline and route attributes. Furthermore, in his in-depth discussion of why airlines do not compete on safety, Savage (2011) stated that airlines would only offer differential safety levels if there is heterogeneity among passengers in the trade-off between safety and price, but he stated that he was not aware of any research on this. Our results make clear that indeed considerable heterogeneity exists, which partly can be attributed to a psychological trait, namely risk taking attitude: passengers with a higher risk taking attitude prefer paying a lower ticket price and accept lower safety levels.

4. Conclusions and discussion

The aim of this paper was to develop and apply a data collection and modeling approach to estimate air passengers' willingness to pay values for safety improvements in air passenger transport. The developed approach is based on the assumption that air passengers first evaluate airline and route characteristics in terms of perceived safety, and then trade-off this safety perception value with regular flight attributes such as travel costs. Accordingly, two stated preference experiments have been developed that are linked together by a safety perception scale. In the first experiment, combinations of airline and route attributes are rated on a safety perception scale. From these observed ratings an ordinary regression model is estimated that indicates to what extent each attribute determines safety perception. In the second experiment, safety perception is included as an attribute in the stated choice experiment, together with the attributes travel costs, travel time and comfort. From the choices observed in this experiment, a Panel Mixed Logit model is estimated which indicates how safety perception is traded-off against the other attributes. This modeling approach allows calculating the willingness to pay values for improvements made in airline and route attributes.

This approach revealed plausible results which can be summarized as follows. The model estimated from the perceived safety experiment made clear that objective airline safety index has a significant impact on the safety perception of a flight. Although such an objective airline safety index for airlines does not exist, the results suggest that such an index would be highly valued by younger passengers, while older passengers apparently base their perception of safety on other attributes. Safety perception is also substantially affected by airlines' involvement in accidents with fatalities. It is important to note that air passengers perceive full service carriers to be more safe than low cost carriers, which is especially the case for those passengers that have been involved in (near) airplane accidents. The results further indicate that flying in bad weather conditions has a substantial effect on safety perception, as has flying over conflict areas, the latter especially applies to older passengers. Conversely, flying over water has only a limited impact. The model estimated from the flight choice experiment revealed that safety perception substantially influences flight choices;

however, the impact decreases with higher values of perceived safety, as one would expect. Furthermore, we found significant interaction effects of the cost parameter with income and risk-taking attitude and the results suggest heterogeneity in the tastes for travel costs, comfort and perceived safety, and no heterogeneity with respect to travel time.

Based on the trade-offs between safety perception and travel cost as observed in the choice experiments, the willingness to pay values for safety improvements were calculated. An illustration of WtP values for the median income and risk-taking travel class based on the median of the WtP values for all statistically significant airline and route attributes is provided. Savage (2011) argues that airlines do not compete on safety and use this in their marketing strategy, because they are afraid that this puts too much attention on the safety issue, which may decrease air travel demand in general. We cannot comment on this statement based on the results of our research, but our results suggest that safety considerations play a substantial role in flight choices and air passengers are willing to pay large amounts of money for flying with higher classified airlines. Our results may be used as a plea to introduce an objective airline safety index for the air industry, analogous to the car industry. This may stimulate airlines to improves their safety procedures in order to obtain a higher objective safety ranking.

Notwithstanding, it should be kept in mind that the presented willingness to pay values are based on a relatively small convenience sample of air passengers recruited in the Netherlands only. Although all categories of the background variables were well presented and we have no indications that a specific sample was realized, the results should be treated with care. On the plus side, we found that our value of time estimate for saving one hour of air travel time was fairly similar to the most recent official Dutch value of time estimate for air travel, which gives confidence in the results of our experiment. Nevertheless, repeating this research in larger, more representative samples and in different countries is advisable.

In addition, the modelling approach needs some discussion, which also offers some directions for further research. First, the hierarchical information integration approach that inspired our modeling approach is not a widely adopted approach and its validity has not been sufficiently researched. This approach assumes that passengers first evaluate airline and route alternatives to arrive at safety perception, which is then traded-off against readily available flight attributes as travels costs. As the experimental set-up follows this assumption, the airline and route alternatives are not directly traded-off against travel costs and other flight attributes. The question therefore is whether the same trade-off values and consequently the same willingness to pay values would be derived if these attributes would be directly included in the choice experiment. However, because including all airplane and route alternatives in the choice experiment would probably result in information overload for the respondents, which is the very reason we adopted this approach, this can only be tested with a limited number of attributes. If such a test indeed shows that similar values can be obtained, then the modeling approach as developed and applied in this study offers large benefits as it allows to include an otherwise prohibitively large number of attributes. We included six airline and route alternatives in this study, but potentially more could be included.

A second remark is related to the information provided in the experiment. By presenting the airplane and route alternatives, we estimate their effect within the context for which this information is known by the air passengers. In reality, this information is not readily available and it depends on the motivation, skills, experience and knowledge of the passenger to obtain that information. For example, knowledge about geography is required to assess information about flight routes. It is not known to what extent air passengers are familiar with this information or actively

acquire this information. If this is true only to a limited extent, then the impact of airline and route attributes on flight choices might be overestimated. Hence, further research on knowledge and motivation to search for safety information is needed.

A third remark can be made with respect to the assumed utility function. Like most logit model applications, the estimated model assumes an additive utility function. This implies that low safety levels can (at least partially) be compensated by high levels of other attributes. While such an additive utility function makes sense in most applications, the question is whether this also applies for safety considerations or whether non-compensatory behavior plays a stronger role in this case. To examine this, other model types that allow for non-compensatory choice behavior may be explored (e.g., Chorus, 2014; Leong and Hensher, 2012; Sælensminde, 2006).

A final remark is related to the adopted 7-point safety perception rating scale. We earlier discussed that one may argue that this is an ordinal scale, whereas we treated this as an interval scale in order to allow for interpolation. In hindsight, it would probably have been better to use a 5-point scale and select all five levels as the levels of the safety attribute in the choice experiment. This allows for representing all levels by dummy variables and avoid the need for interpolation. Assuming interval measurement level would then not be required and an ordered logit model could be estimated from the safety perception experiment instead of an ordinary regression model. To some extent, this also solves another potential problem, which is that not all respondents used the entire range in the rating experiment. Some respondents only gave low values, while others gave only high values, reflecting how safe they generally perceive the presented airline and route attributes combinations to be. However, also these respondents were presented with the entire range of safety perception values in the choice experiment. It is not clear what these respondents assumed about which airline and route attributes would be associated with values they may perceive as extreme. Hence, reducing the range of the rating scale may reduce this potential problem. Should this not be sufficient, than an alternative approach could be adopted, which is to customize the choice experiment. This would allow for the presentation of only the levels of the safety perception attribute that are within the range of values given by each respondent in the rating experiment.

To conclude, although the results of the applied modeling approach seem plausible, still some methodological research may be required to further increase confidence in the generic applicability of obtained results.

References

Ben-Akiva, M.E., Ramming, M.S. and Walker, J.L., 1998. Improving behavioural realism of urban transportation models through explicit treatment of individuals' spatial ability. Paper presented at Transportation Models in the Policy-Making Process: A Symposium in Memory of Greig Harvey, Asilomar Conference Center, CA, USA.

Bierlaire, M., 2016. PythonBiogeme: a short introduction. Report TRANSP-OR 160706, Series on Biogeme. Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Bliemer, M.C.J., Rose, J.M., 2011. Experimental design influences on stated choice outputs. An empirical study in air travel choice. Transportation Research Part A, 45(1), 63-79.

Boksberger, P.E., Bieger, T., Laessar, C., 2007. Multidimensional analysis of perceived risk in commercial air travel. Journal of Air Transport Management. 13(2), 90–96.

Bonsall, P., Firmin, P., Beale, J., 2004. Perception of modal attributes: how accurate and how susceptible to change? Paper presented at the 83rd Meeting of the Transportation Research Board, Washington DC.

Bos, D.M., Heijden, R.E.C.M Van der., Molin, E.J.E., Timmermans, H.J.P., 2004. The choice of park & ride facilities: an analysis using a context-dependent hierarchical choice experiment, Environment and Planning A, 36(9), pp. 1673–1686.

Brey, R., Walker, J.L., 2011. Latent temporal preferences: an application to airline travel. Transport. Transportation Research, Part A, 45 (9), 880–895.

Brownstone, D., Small, K.A., 2005. Valuing time and reliability: assessing the evidence from road pricing demonstrations. Transportation Research Part A: Policy and Practice, 39(4), 279-293.

Chang, Y.-H., Yeh, C.-H., 2004. A New Airline Safety Index. Transportation Research Part B 38, 369-383.

Chorus, C.G., 2014. Capturing alternative decision rules in travel choice models: A critical discussion. Chapter 13, 290-310. In Hess, S. & Daly, A. (Eds.) Handbook of Choice Modelling. Edward Elgar

Collins, A., Rose, J.M., Hess, S., 2012. Interactive stated choice surveys: a study of air travel behaviour. Transportation, 39, 55–79.

Fleischer, A., Tchetchik, A., Toledo, T., 2012. The Impact of Fear of Flying on Travelers' Flight Choice: Choice Model with Latent Variables. Journal of Travel Research, 51 (5), 653-663.

Fleischer, A., Tchetchik, A., Toledo, T., 2015. Does it pay to reveal safety information? The effect of safety information on flight choice. Transportation Research, Part C, 56, 210–220.

Foreman, E., Gerwen, L. Van, 2008. Fly Away Fear: Overcoming Your Fear of Flying. London: Karnac Books Ltd.

Golledge, R.G., 2002. Dynamics and ITS: behavioural response to information available from ATIS, in: H. S. Mahmassani (Ed.) In: Perpetual Motion: Travel Behaviour Research Opportunities and Application Challenges (Amsterdam: Pergamon).

Hall, M.C., 2002. Travel safety, terrorism and the media: the significance of the issue-attention cycle. Current Issues in Tourism, 5, 458–466.

Hess, S., Adler, T., Polak, J.W., 2007. Modelling Airport and Airline Choice Behaviour with the Use of Stated Preference Survey Data. Transportation Research Part E, 43, 221-233.

IATA, Safety Report 2015 (2016). accessed 18 Okt 2016.

Kasperson, R.E, Renn, O., Slovic, P., Brown, H.S. Emel,J. Goble, R., Kasperson, J.X., Ratick, S. 1988., The social Amplification of Risk, Risk Analysis, 8(2), 177-187.

Kebabjian, R., 2015. (March 15). Airline Accident Ratings. Retrieved from Plane Crash Information: http://www.planecrashinfo.com/accidents.htm

KIM (Kennisinstituut Mobiliteit Nederland) (2013), Value of time and reliability in passenger and freight transport in the Netherlands, Report for the Ministry of Infrastructure and the Environment, http://www.kimnet.nl/publicaties/rapporten/2013/11/18/values-of-time-and-reliability-in-passenger-and-freight-transport-in-the-netherlands (accessed, July 29 2016).

Koo, T.T.R, Caponecchia, C., Williamson, A., 2016. Measuring the effect of aviation safety risk eduction on flight choice in young travellers. Safety Science, 73, 1-7.

Koo, T.T.R, Caponecchia, C., Williamson, A., 2016. How important is safety in making flight choices? Evidence from simple choice experiments, Transportation, on-line since 15 September 2016.

Koppelman, F.S., Pas, E.I., 1980. Travel choice behaviour: models of perceptions, feelings, preference, and choice, Transportation Research Record,765, 26–33.

Leong, W., Hensher, D.A., 2012. Embedding decision heuristics in discrete choice models: A review. Transport Reviews, 32(3), 313-331.

Louviere, J.J., 1984. Hierarchical Information Integration: a new method for the design and analysis of complex multi-attribute judgement problems, in: T. C. Kinnear (Ed.) Advances in Consumer Research, 11, 148–155 (Provo, UT: Association for Consumer Research).

Molin, E.J.E., Timmermans, H.J.P., 2009. Hierarchical Information Integration Experiments and Integrated Choice Experiments, Special Issue on Observing Complex Choice Behavior with Stated Preference Experiments: Innovations in Design, Transport Reviews, 29(5), 635-655.

Molin, E.J.E., Gelder, M. van, 2008. Freeway access to public transport: a hierarchical multimodal choice model, Transportation Research Records, 2076, 106–113.

Oppewal, H., Louviere, J.J. and Timmermans, H.J.P., 1994. Modeling hierarchical conjoint processes with integrated choice experiments, Journal of Marketing Research, 31, 92–105.

Rhoades, D.L., Waguespack, B., 1999. Better safe than service? The relationship between service and safety quality in the US airline industry. Managing Service Quality: An International Journal, 9 (6), 396–401.

Rhoades, D.L. and Waguespack, B., 2000, Judging a book by its cover: the relationship between service and safety quality in U.S. national and regional airlines. Journal of Air Transport Management, 6, 87-94.

Rose, N., 1992. Fear of Flying–economic analyses of airline safety. Journal of Economic Perspectives, 6(2), 75–94.

Savage, I., 2011. Competition on the basis of safety? In: James Peoples (Ed.), Pricing Behavior and Non-Price Characteristics in the Airline Industry Advances in Airline Economics, vol. 3, Emerald Group Publishing Limited, 297–323.

Sælensminde, K., 2006. Causes and consequences of lexicographic choices in stated choice studies. Ecological Economics, 59(3), 331-340.

Squalli, J., Saad, M., 2006. Accidents airline safety perceptions and consumer demand. Journal of Economics and Finance, 30(3), 297–305.

Thorpe, J., 2003. Fatalities and Destroyed Civil Aircraft due to Bird Strikes 1912-2002. Warsaw: International Bird Strike Committee.

Weber, E.U., Blais, A.-R., Betz, N.E., 2002. A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviors. Journal of Behavioral Decision Making, 263-290.

Wen, C.H., Lai, S.C., 2010. Latent class models of international air carrier choice. Transport. Res. Part E – Logistics and Transportation Review, 46 (2), 211–221.