

# Abstracts

## Contents:

### Austria

Eva Benková  
Simona Jokubauskaite  
Sarah Karasek

### Croatia

Iva Budimir  
Ivana Lolić  
Ivan Papić

### Hungary

Bálint Nagy  
Laszló Németh  
György Márk Kis

### Italy

Teresa Barrella  
Daniele Bottigliengo  
Veronica Sciannameo

### Slovenia

Marjan Cugmas  
Daria Maltseva  
Ana Slavec  
Joca Zurc

## Austria

### Privacy Sets for Constrained Space-Filling

**Eva Benková (Austria)**, Radoslav Harman, Werner Müller

e-mail: [eva.benkova@jku.at](mailto:eva.benkova@jku.at)

Optimal design of experiments studies methods of conducting an experiment so that the maximum amount of information on the unknown parameters is attained. Besides natural restrictions on time or cost of the experiment, it is often necessary to ensure space-fillingness of the design, i.e., an appropriate distribution of the measurements across the design space.

This can be done by what we call “soft” methods, i.e., by optimizing a prescribed objective measuring a degree of space-fillingness, such as maximin, minimax, etc. In contrast, hard space-filling methods ensure desirable properties by enforcing constraints on the designs, such that a secondary criterion (e.g. D-optimality) can be used for optimization. We propose a general framework for the latter methods based on the central notion of so-called privacy sets. This notion provides a unifying framework for standard designs without replication, Latin hypercube designs, and Bridge designs, among many others.

We introduce a heuristic exchange-type algorithm for the construction of good information designs under conditions expressed by a system of privacy sets and compare its performance on some well-known examples. For instance, we demonstrate that for the computation of Bridge designs this algorithm performs significantly better than the state-of-the-art method. Moreover, the application of privacy sets is not restricted to cuboid design spaces and promises improvements for many other situations.

### Joint estimation of Discrete Choice and Time-Expenditure Allocation System

**Simona Jokubauskaite (Austria)**, Reinhard Hössinger, Florian Aschauer, Friedrich Leisch,

Regine Gerike, Sergio Jara-Diaz, Stefanie Peer, Basil Schmid, Kay W. Axhausen

e-mail: [simona.jokubauskaite@boku.ac.at](mailto:simona.jokubauskaite@boku.ac.at)

The value of travel time savings (VTTS) represents the willingness to reduce travel time in favor of other activities that generate more utility. It is usually gained from indirect utility models estimating the willingness to reduce travel time in the context of mode choice, route choice, or destination choice. The VTTS has always been subject to extensive debate in both academia and politics, because savings in travel time are the major justification for transport infrastructure investments. Recent research has shown a trend away from pure VTTS to a more informative measures: the value of leisure (VOL) in terms of the marginal willingness-to-pay for an additional unit of leisure, the value of time assigned to work (VTAW), and the value of time assigned to travel (VTAT), which is derived from deducting the VTTS from the VOL. State of the art models claim that a person who makes a travel decision not only maximizes his utility in this particular choice, but also in the time-and-expenditure allocation space. In

this paper we present for the first time a joint time-and-expenditure allocation and discrete mode choice model, which includes all required components. Model estimation is based on full information maximum likelihood method. We use a novel dataset including all activities, expenditures and travel decisions over a period of one week, all of which were collected simultaneously, i.e., from the same individuals at the same time. The net sample includes 748 representatively selected Austrian workers. The model estimates the VOL to be approximately 8.9 €/h and the VTAW about -3.2 €/h. It implies that work is displeasure and people work mainly for the wage (the sample mean of the wage is 12.1 €/h). The VTTS differs substantially between travel modes. It ranges from 3.8 €/h for public transport and 9.6 €/h for car to almost 13 €/h for walking.

## **Model-Based Clustering to Classify Wafer Images**

**Sarah Karasek** (Austria), Herwig Friedl, Peter Scheibelhofer  
e-mail: [s.karasek@tugraz.at](mailto:s.karasek@tugraz.at)

In the semiconductor fabrication frequent quality checks during wafer production are essential. For the quality assessment of the so-called wafer bond process, one can take a look at the grayscale image of a wafer, which looks like a black-and-white photograph. Before the quality of the wafer can be judged by analyzing the grayscale image, it is important to know if the image contrast of the picture is high enough, because only then a reliable assessment is possible.

This can be done by considering the grayscale frequencies. The corresponding histograms show up a clear multi-modal behaviour. That is why we assume a finite mixture distribution with component specific parameters to analyze the data. In some cases, a mixture of Gaussians works very well, whereas sometimes a mixture of Gammas performs even better. In order to find the maximum likelihood estimates of the location and scale or shape parameters we utilize the well-known EM algorithm. Once we have estimates for a set of images that have been already categorized by an expert, we use specific characteristics of their component distributions to correctly assign these images to their quality levels.

## **Croatia**

### **Effects of Normalization Procedures to Statistical Inference in Omics Research**

**Iva Budimir** (Croatia)  
e-mail: [ibudimir@gradst.hr](mailto:ibudimir@gradst.hr)

A common problem in modern Omics research is that measurements are affected by complicated experimental procedures leading to differences between replicates and the existence of batch effects, among others. This has been observed in the recent years in fields like metabolomics, metagenomics, proteomics, glycomics and transcriptomics, among others. One of the most used approaches to reduce variability between replicates and to improve comparisons between samples is to normalize the data. Normalization is the attempt to

compensate for systematic technical differences between measurements thus allowing researchers to assess real biological differences between samples. However, since normalization procedures transform the data subsequent statistical inference procedures performed on the data might be biased by the underlying normalization or not usable at all. In this talk we will give a brief introduction into some of the most widely used normalization procedures and present some of the results we have obtained on the effects of normalization techniques in glycomics research to regression modelling and variable selection.

*Keywords:* omics, normalization, noise, inference, regression

*AMS subject classifications:* 62P10, 92C40

## **Estimating Inflation Expectations from Consumer Surveys**

**Ivana Lolić** (Croatia)

e-mail: [ilolic@efzg.hr](mailto:ilolic@efzg.hr)

Maintaining price stability is the most important assignment of a central bank. During the Great Recession, most central banks decreased interest rates to the historically lowest values, all to stimulate economic growth and prevent stagnation. However, despite all efforts, most developed countries have recorded a long period of deflation and economic slowdown. So, various questions about the monetary policy were raised: why did the decrease in interest rates not prevent deflation, or why did it last so long? This has once again highlighted the importance of consumers' perceptions and expectations of inflation. Even though policyholders often forget that peoples' opinions do matter. However, extracting assessments about inflation is a difficult task. Harmonized surveys for consumers comes as a natural tool for quantifying them, which is why the questions about inflation perceptions and expectations constitute a significant part of the European Business and Consumer Surveys. The focus of this paper is the probabilistic quantification method. The main idea is to fit the observed answers' frequencies to a probabilistic distribution (such as normal, Student's, etc.). However, the main disputes are led over the "correct" distribution and the set of imposed restrictions (such as the long-term unbiasedness of expectations, time-constant intervals, etc.). This paper provides a thorough quantitative investigation, which looks at more than 2500 estimation specifications (different combination of distributions and restrictions). The analysis provides conclusions about the "best" estimates of inflation expectations, the price tag of distribution assumption, and the cost of every utilized restriction. This study also provides a novel set of assumptions that are the least restrictive and incorporate the European Central Bank's inflation targeting policy.

*Keywords:* Business and Consumer Surveys, Inflation Expectations, Inflation Perceptions, Carlson and Parkin method

## Correlated Continuous Time Random Walks and Pearson Diffusions

Ivan Papić (Croatia), Nikolai N. Leonenko, Alla Sikorskii, Nenad Šuvak  
e-mail: [ipapic@mathos.hr](mailto:ipapic@mathos.hr)

Continuous time random walks have random waiting times between particle jumps. We define the correlated continuous time random walks (CTRWs) that converge to Pearson diffusions. Pearson diffusions have stationary distributions of Pearson type (i.e. normal, gamma, beta, Fisher-Snedecor, reciprocal gamma or Student stationary distribution). The jumps in these CTRWs are obtained from Markov chains through the Bernoulli urn-scheme model and Wright-Fisher model. The jumps are correlated so that the limiting processes are not Lévy but diffusion processes with non-independent increments. According to the carefully chosen starting Markov chain, different limiting Pearson diffusions can be obtained. We present Ornstein-Uhlenbeck (OU) process, Cox-Ingersoll-Ross (CIR) and Jacobi diffusion as the limiting Pearson diffusions.

*Keywords:* Markov chains; diffusions; Pearson diffusions; urn-scheme models; Wright-Fisher model; continuous time random walks

## Hungary

### Assessing Robustness of Estimated Treatment Effect Using Sensitivity Analysis

Bálint Nagy (Hungary)  
e-mail: [balint.nagy.tnv@gmail.com](mailto:balint.nagy.tnv@gmail.com)

Propensity score matching is a popular method to estimate average treatment effect on the treated in observational studies. However, checking the most important assumption of this method is often omitted by its users: the final estimate is unbiased only in that case if the matching model includes all covariates that affect the process of allocation into treatment and control groups. For this reason researchers using PSM need to perform a sensitivity analysis to assess the validity of their estimates.

A sensitivity analysis helps to decide whether including a hypothetical unobserved covariate in the matching model would significantly alter the estimate. If the estimate is robust to the inclusion of many different unobserved covariates it could be considered as valid.

In my research I analyzed data from the National Assessment of Basic Competencies in Hungary conducted in 2014 to answer the following question: do independent private school students perform better at math and reading than state school students? According to the result of propensity score matching there is no significant difference between the performances of the two groups. However, the results of two different types of sensitivity analysis suggest this estimate can be strongly biased due to unobserved covariates.

# Distribution Dependence of Tail Index Estimators

**Laszlo Németh** (Hungary)

e-mail: [laszlo.nemeth.92@gmail.com](mailto:laszlo.nemeth.92@gmail.com)

In extreme value theory the rare and high values of a distribution have big influence. These extreme values of a distribution can be measured by the tail index parameter. One of the widely used estimators for the tail index is the Hill estimator, which is based on the  $k$  upper order statistic. Finding the optimal  $k$  is aimed by recent works by using the double bootstrap [1], or the Kolmogorov-Smirnov metric[2]. These methods are asymptotically consistent, but the application were tested only for  $\xi < 0.5$

I realised that the method based on Kolmogorov-Smirnov distance results in biased estimators when the tail index  $\xi > 0.5$  and the sample size is less than 10 000. Moreover the distribution of the estimator can be modelled by a generalized extreme value (GEV) distribution. I observed a linear relation between the location parameter of the fitted GEV distribution and the theoretical tail index in case of ( $0.5 < \xi < 4$ ). Using this linear connection and the  $m$  out of  $n$  bootstrap technique I constructed a new method for tail index estimation, named as fitted regression method.

Generally the fitted regression method performs better than the one, based on Kolmogorov-Smirnov distance, and is similar to the double bootstrap in the interval ( $0.5 < \xi < 4$ ). For different initial distributions I compared the slow but precise double bootstrap method to the fitted regression method. This way one can choose the better model for each distribution. I have developed a model selection method for real life data as well.

## *References*

1. J. Danielsson, L. de Haan, L. Peng, C. G. de Vries, Using a bootstrap method to choose the sample fraction in tail index estimation, *Journal of Multivariate Analysis*, Volume 76, Issue 2, February 2001, pp. 226-248, 2001.
2. J. Danielsson, L. M. Ergun, L. De Haan, C.G. de Vries, Tail Index Estimation: Quantile Driven Threshold Selection, manuscript [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2717478](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2717478), 2016

## **Finding the Playmaker - Key Player Identification in American Football**

**György Márk Kis** (Hungary)

e-mail: [kis.gyorgy.mark@gmail.com](mailto:kis.gyorgy.mark@gmail.com)

The paper takes an unorthodox view on team sports: it aims to analyze American Football as a form of networks, and find the key player based on methods used by Ballester et al. (2006) in crime networks. Utility of players and their actions was based on differences in actual and expected contributions to the game flow and scoring, thus assigning point values to every on-field action. The results of the analysis of the 2015 Minnesota Vikings defense using the concepts outlined showed clear signs of usefulness given automation, but also provided an insight into the mechanics of said team. Shortcomings of the method and their possible solutions were also discussed as a basis for improvement.

## **Italy**

### **GeCUB Models for a Subjective Evaluation of Functional Septoplasty**

**Teresa Barrella** (Italy)

e-mail: [terryb86@tiscali.it](mailto:terryb86@tiscali.it)

*Aim of the study.* The patient's quality of life has become a new parameter, alongside the medical/clinical indicators used for assessment of the outcome of treatments. Hence the object of the present study is to detect the subjective perception of improvement of nasal breathing following surgery of functional septoplasty, and to provide a kind of "qualitative" measure of the success of the intervention same. In particular the novelty of the study is the proposal of an alternative model for ordinal data, CUB models (Piccolo, 2003), which aims to analyze the psychological mechanism that drives the individuals' judgment in rating surveys; an approach based on a discrete mixture, interpreting the final response of subjects as a weighted combination of a personal "feeling" and some intrinsic "uncertainty".

*Materials and methods.* To evaluate the recovery of the nasal functionality, two different questionnaires have been "ad hoc" prepared and administered to patients: one in the pre-operative phase, the other one to 7 days after surgery. Data were collected at the Department of Otolaryngology at the University Studies of Campania "Luigi Vanvitelli", from March to December 2016.

In the present case the most recent generalization of CUB models has found application, which adds a third element to the mixture, the "shelter effect", supposing that in some circumstances a subgroup of respondents chooses a specific option only to simplify a more demanding choice. Through this generalization, known as GeCUB models (Iannario, Piccolo, 2016), that allows to study the effect of subjects' covariates on all three components of the mixture, has been analyzed the relationship between the global post-surgery satisfaction and socio-demographic and biometrics features of the patients, and the same overall evaluation expressed in the pre-surgery phase by them.

*Results.* The sample obtained consists of 150 patients, mainly males (77%), of average age at around 35. Most of the patients felt more difficulty breathing during physical activities and at night. In post-surgery, 93% of the sample breathes better and there are significant improvements in the symptoms associated with the disease (roncopathy, compensatory oral respiration, hyposmy). After the estimation of different CUB models, a CUB model with “shelter effect” at the extreme category rating scale (corresponding to the maximum satisfaction) was selected, given a better goodness of fit. Thanks to the introduction of subjects’ covariates, the empirical analysis showed a significant relationship between the level of overall satisfaction after surgery and the type of work performed, the age, the weight and the overall level of satisfaction expressed in the pre-operative phase.

*Conclusions.* It can be concluded that in the surveys where the respondents' opinion is measured with rating scales, the proposed model is a great alternative to the classical approach, both in statistical terms (significance, fitting, comparisons), and in terms of interpretation and graphic representation. In particular with the introduction of a GeCUB model, it was possible to grasp an aspect not detectable by a traditional model for ordinal data, i.e. the relationship between the evaluation before and after the surgery, from which it is assumed that the more patients felt respiratory discomfort in the first phase, the more they had a perception of improvement in post-surgery phase.

## **Evaluation of Hadrontherapy Efficacy: a Bayesian Approach Combining Clinical and In-Vitro Data**

**Daniele Bottigliengo** (Italy), Andrea Attili, Paola Berchiolla  
e-mail: [daniele.bottigliengo@unito.it](mailto:daniele.bottigliengo@unito.it)

In the last few decades, a growing interest raised toward hadrontherapy, a new frontier of radiotherapy that possess some attractive features that make it more efficient in cancer treatment respect to traditional radiotherapy. One of the major research topic in hadrontherapy is focused on the effect produced by particle beams on irradiate cellular tissues. This effect is assessed by the Relative Biological Effectiveness. The aim of this study is to provide a new Relative Biological Effectiveness estimation method taking advantage of the Bayesian approach.

Two sources of data were employed during the analysis: one dataset containing *in-vitro* experiments and one dataset containing clinical data generated from parameter values of 6 individuals treated for prostate cancer. Relative. The former one was used to elicit the *a-priori* distribution to enter in a Bayesian model for deriving the posterior distribution of the parameters of the radiobiological models that are used to assess the biological effect linked to the clinical outcome of ion beam treatment.

Despite the small sample size of clinical data it was possible to estimate the Clinical Relative Biological Effectiveness and the related uncertainty. This information can be used as support for planning personalized treatment.

# Semi-Bayesian Method for Occupational Studies: An Example from a Pooled Analysis of Two Italian Case-Control Studies on Bladder Cancer

**Veronica Sciannameo** (Italy), Angela Carta, Angelo d'Errico, Maria Teresa Giraud, Francesca Fasanelli, Cecilia Arici, Milena Maule, Paolo Vineis, Stefano Porru, Carlotta Sacerdote, Fulvio Ricceri  
e-mail: [veronica.sciannameo@unito.it](mailto:veronica.sciannameo@unito.it)

*Introduction.* The main aim of occupational epidemiology is to identify health risks due to occupations. Occupational exposures are usually studied in occupational cohorts, but they could be studied also in other types of study (case-control or cohort studies), considering roughly the subjects' jobs. However, this latter approach faces the problem of multiple comparisons that increases false positive results and the problem of scarcity of data (few subjects for each occupation). Traditional approaches do not allow including *a priori* information about occupational risks in the analyses and Bayesian methods could be helpful in this situation. Among cancers, bladder cancer (BC) is interesting to study in this framework because the main causal risk factors are cigarette smoking and exposure to occupational carcinogens.

*Objectives.* The main objective of this study is to apply a Bayesian approach to analyze the association between BC risk and exposure to certain jobs, controlling for spurious results and considering *a priori* information. Second, the associations between BC and occupational exposure to selected carcinogen agents will be analyzed, also separating by tumor grade.

*Materials and methods.* We pool data from two hospital case-control studies on male BC conducted in the North of Italy (Turin and Brescia). A detailed questionnaire about life-style, including smoking habits and working history, was compiled by 893 cases and 978 controls. Occupations are classified by means of the International Standard Classification of Occupations (ISCO-68). Exposure to carcinogens is identified using a validated Job Exposure Matrix. The main analyses are performed by means of a two-stages hierarchical model: first, standard multivariate logistic regressions are conducted; then, a semi-Bayesian approach is applied. Briefly, in the latter approach, the value of the variation of the true risk is *a priori* decided. We set it to 0.23, 0.41, 0.59 and 0.76, assuming that 95% of relative risks are within a 2.5, 5, 10 and 20 fold-range of each other respectively, according to previous works in literature. Finally, in the analyses on carcinogen agents, multivariate logistic regression models are constructed.

*Results.* A significantly increased BC risk is found in the first stage of the hierarchical model for chemical engineering technicians (OR=12.67, 95% CI 1.47-109.24), postmen (OR= 4.99, 95% CI 1.05-23.66), and for lathe operators (OR=1.67, 95% CI 1.05-2.67). This last result remains significant after Bayesian control and sensitivity analysis. In the analyses on carcinogen agents, cadmium and trichloroethylene are found to be significantly associated with BC risk (respectively, OR=1.24, 95% CI 1.02-1.52 and OR=1.23, 95% CI 1.00-1.51). Finally, separating by tumor grade, carcinogens seem to be more effective in low-grade BC.

*Conclusion.* The two-stages hierarchical Bayesian model seems to be efficient, allowing to control for false positive risk using *a priori* information, and to obtain more realistic occupational risk estimates. This study confirms already known associations of jobs and carcinogen agents with BC risk, but for the first time is found a greater effect in low-grade BC.

## **Slovenia**

### **Generating Networks with a Given Blockmodel by Different Triad Types**

**Marjan Cugmas** (Slovenia), Anuška Ferligoj, Aleš Žiberna  
e-mail: [marjan.cugmas@fdv.uni-lj.si](mailto:marjan.cugmas@fdv.uni-lj.si)

The presentation discusses whether one can generate networks with a given blockmodel (i.e., cohesive, core-periphery, hierarchical, and transitivity), considering only different triad types. The inability to generate such networks can be the consequence of the characteristics of the algorithm for generating networks. Therefore, two different algorithms are applied: (i) the method of relocating links and (ii) the less deterministic Monte Carlo Multi Chain algorithm implemented in the ergm package in R.

Using one or another algorithm, one can generate the networks with a given blockmodel, considering only different triad types. Furthermore, considering only a subset of triad types it is possible to improve the expected blockmodel type of a network. Considering some other local network structures can substantially improve the expected blockmodel type of a network, e.g., considering paths of length three in the case of a hierarchical blockmodel with complete blocks on the diagonal.

### **Studying the Community Structure of Network Researchers In Russia**

**Daria Maltseva** (Slovenia)  
e-mail: [d\\_malceva@mail.ru](mailto:d_malceva@mail.ru)

The direction of network studies in Russia is a new form of a research methodology. However, in recent years we can observe the growing number of scientists who implement the instruments of Social Network Analysis into their studies in different disciplines, such as sociology, economy, political studies, management, etc. The procedure of mapping this emerging community of Russian network researchers is not only important from epistemological point of view, but also has practical meaning as it can show the main active clusters and main groups of knowledge exchange – that could help to increase the possibility of cooperation between scientists.

We study structure of the research community of Russian scientists involved into network research by the analysis of articles devoted to network topics published in Russian scientific journals. We use data from the electronic library of Russian-language scientific periodicals eLibrary.ru, which contains more than 3800 journals. On the first step, using the key words corresponding to network studies, the data set of more than 5 000 articles was collected. On the second step, we collect all the references that were mentioned in the bibliography lists at the initial articles.

Data analysis allows us to look both at the thematic agenda of network studies in Russia and the patterns of scientists` cooperation in the field – i.e. collaboration between researchers

from different organizations with each other and with foreign scientists. Presenting the data on the scientific field's development, this study contributes to the studies in the field of sociology of science, as well as to the studies focused on the SNA development in different national scientific traditions.

## **How to Detect Unfamiliar Wordings in Survey Questions: Comparison of Current Questionnaire Pre-Testing Methods and Linguistic Resources**

**Ana Slavec** (Slovenia), Vasja Vehovar  
e-mail: [ana.slavec@innorenew.eu](mailto:ana.slavec@innorenew.eu)

A survey item can be worded in numerous ways, some interchangeable and some not, and it is often difficult to decide which words to select in a specific context. Even wordings that are synonymous might be different in terms of familiarity to the respondent – some terms are more common (familiar), while others are rarer (unfamiliar) and possibly more difficult to comprehend. To some extent, unfamiliar terms can be detected by current survey pre-testing methods. For instance, participants in cognitive interviews can point out comprehensibility issues; however, unfamiliarity is relative and it is important to implement them on wide numbers and different kinds of participants. This can often be expensive. Another pre-testing method that can be used to detect unfamiliar terms and other comprehensibility issues, are expert reviews. However, experts, as well as questionnaire designers, are usually educated people with a wider vocabulary and a tendency to use overcomplicated language. This might make it hard to look from the perspective of a respondent with a narrower vocabulary. In this paper we discuss the potential of linguistic resources such as text corpora and lexical databases to be used in the questionnaire development process to prevent comprehensibility issues. We evaluated wording frequencies based on text corpora as estimates of question familiarity which could be integrated in survey questionnaire tools. In addition, we use wordnets, large lexical databases where words are grouped into sets of cognitive synonyms, to search for suggestions of alternative wordings. Based on our research on three case studies (three questionnaires), we are developing a pre-testing procedure that is able to detect unfamiliar wordings and suggest improvements in survey question wording. The procedure could potentially be developed as a module to be integrated in survey software and used in the question development process in combination with other pre-testing methods.

## Development of core validity assessment criteria for the mixed methods research

**Joca Zurc** (Slovenia), Anuška Ferligoj  
e-mail: [joca.zurc@guest.arnes.si](mailto:joca.zurc@guest.arnes.si)

The mixed methods methodology has emerged and significantly developed over the past 30 years and became recognised as third methodological group in science (Creswell, 2014; Teddlie and Tashakkori, 2009; Koller and Sinitsa, 2009; Johnson and Onwuegbuzie, 2004). Research quality and, in particular, validity of mixed methods research constitutes an extremely important field due to a combination of two different paradigmatic philosophical perspectives, methodologies and research methods in the same study. Therefore, mixed methods scholars emphasize the need for selecting and reaching a consensus on a set of core validity criteria. These core validity assessment criteria can provide guidelines for implementing quality mixed methods research, contribute to a common methodological language, and help different stakeholders such as researchers, funders, reviewers, supervisors, policy makers, study participants, and those interested in study findings. Accordingly, we designed an empirical study with the purpose to determine the mixed methods validity criteria that are mandatory for every mixed methods study and to develop a measurement instrument for self-assessment of validity, which would be useful in planning, implementing and evaluating mixed methods research. The study was made in three sequential empirical research phases following ten interactive steps of IDCV methodological framework—Instrument Development and Construct Validation (Onwuegbuzie et al. 2010), using NVivo and statistical analysis. In the first research phase, a systematic literature review with qualitative meta-synthesis approach was performed to identify which validity assessment criteria have previously been developed and applied in mixed methods research. These findings were used as the basis for developing the qualitative self-structured in-depth interview, which was conducted on a sample of nine mixed methods most distinguished experts who were invited as keynote speakers at the *Mixed Methods International Research Association Inaugural Conference* (2014) held in Boston, USA. In the third research phase, the main focus was on developing and testing the measurement instrument for validity assessment in the mixed methods research. The initial instrument was three times pilot-tested on a sample of 56 undergraduate and graduate students at the Turku University of Applied Sciences in Finland. Considering the findings of three pilot tests, the revised instrument was created and field-tested by three experienced reviewers on ten published mixed methods studies from different scientific fields. Testing of the revised instrument suggested ten core validity assessment criteria for mixed methods research, accepted with consensus: 1) rationales for research question, 2) justification of mixing, 3) mixed methods research design, 4) integration, 5) warranted methodology, 6) three-column approach to validity assessment, 7) warranted findings, 8) society sensitivity, 9) engagement of difference, and 10) believability. The insufficient integration of qualitative and quantitative methods in the selected studies (criterion 4) was the most pronounced. Based on the proposed validity criteria a validity index for mixed methods research can be defined.

*Keywords:* research quality, validity, experts, mixed methods, pragmatism