

Inference from Non Probability Samples

Joint Conference of:
Survey Research Methods (SRM)
European Survey Research Association (ESRA)
Étude Longitudinal par Internet Pour les Sciences
Sociales (ELIPSS)

S

March 16/17 2017

SciencesPo

254 Boulevard Saint-Germain

75007 Paris

France

P

Keynotes:

Jelke Bethlehem (Leiden University, The Netherlands)

Andrew Gelman (Columbia University, New York, USA)

N

Scientific Organizers:

Nick Allum (University of Essex, Chelchester, UK)

Ulrich Kohler (University of Potsdam, Germany)

Laurent Lesnard (Sciences Po, Paris, France)

I

Contents

Schedule	2
Day 1, March 16 th 2017	2
Day 2, March 17 th 2017	3
Abstracts	4
Day 1, March 16 th 2017	4
Day 2, March 17 th 2017	7
Information	13
Conference Venue	13
Registration	13
Map of Venue	14
Public Transport	14
Conference dinner	14
Wifi	14

Schedule

Day 1, March 16th 2017

12:30–13:00 Registration

13:00–13:15 Welcome

Nick Allum, Ulrich Kohler and Laurent Lesnard

Statistical Issues

13:15–13:45 Nonprobability sampling as model construction: Expanding beyond the ideal of randomization

Andrew Mercer

13:45–14:15 A Partially Successful Attempt to Integrate a Web-Recruited Cohort into an Address-Based Sample

Phillip S. Kott, Matthew Farrelly, Kian Kamyab

14:15–14:45 A test of sample matching using a pseudo-web sample

Jack Gambino and Golshid Chatrchi

14:45–15:00 Coffee and Refreshments

15:00–15:30 Expanding the toolbox: inference from non-probability samples using machine learning

Joep Burger, Bart Buelens, Jan van den Brakel

15:30–16:00 Investigation into the use of weighting adjustments for non-probability online panel samples

Dina Neiger, Darren W. Pennay, Andrew C. Ward, Paul J. Lavrakas

16:00–16:30 A bootstrap method for estimating the sampling variation in point estimates from quota samples

Jouni Kuha, Patrick Sturgis

16:30–16:45 Coffee and Refreshments

Keynote

16:45–17:45 Looking for rigor in all the wrong places

Andrew Gelman, Columbia University, New York

Day 2, March 17th 2017

- 8:50–9:00 Welcome
 Nick Allum, Ulrich Kohler and Laurent Lesnard
- Comparison of Probability and Non-Probability Samples*
- 9:00–9:30 The Accuracy of Online Surveys: Evidence from Germany
 Annelies G. Blom, Daniela Ackermann-Piek, Susanne Helmschrott, Carina Cornesse, Christian Bruch, Joseph W. Sakshaug
- 9:30–10:00 Assessing the Accuracy of 51 non-probability online panels and river samples: A re-analysis of the Advertising Research Foundation (ARF) online panel comparison experiment
 Mario Callegaro, Yongwei Yang, Katherine Chin, Ana Villar, Jon A. Krosnick
- Keynote*
- 10:00–11:00 The perils of non-probability sampling
 Jelke Bethlehem
- 11:00–11:15 Coffee and Refreshments
- How to collect non-probability samples*
- 11:15–11:45 An Empirical Process for Using Non-probability Survey for Inference
 Robert Tortora and Ronaldo Iachan (ICF)
- 11:45–12:15 Inbound Call Survey (ICS) – A New Methodology
 Karol Krotki, Burton Levine, Georgiy Bobashev, Scott Richards (RTI and Reconnect Research)
- 12:15–12:45 In search of best practices
 Sander Steijn and Joost Kappelhof (The Netherlands Institute for Social Research/SCP)
- 12:45–13:00 Publication Plans (Special Issue of SRM)
- 13:00 *End of the meeting*

Abstracts

Day 1, March 16th 2017

13:15–13:45

Nonprobability sampling as model construction: Expanding beyond the ideal of randomization

Andrew Mercer

(Pew Research Center, University of Maryland)

AMercer@PewResearch.org

Both in practice and in the methodological literature, there exists a widespread expectation that nonprobability samples should have similar properties to probability-based samples – that researchers should be able to commission a survey using a standard data collection procedure, apply a standard set of demographic quotas or weights, and draw reliable inferences about a wide range of topics. When such samples yield biased estimates, this is taken as evidence that the output of the nonprobability survey process insufficiently mimics the process of random selection. Rather than evaluate nonprobability samples in terms of their resemblance to the probability-based ideal, this paper argues that nonprobability sampling is better viewed as part of model construction, where the researcher must identify confounding variables and specify their distribution explicitly and in advance. This perspective sees the distinction between probability-based and nonprobability survey inference as analogous to the distinction between causal inference from randomized experiments and observational studies. We review how this framework is guiding the Pew Research Center’s ongoing research into the use of nonprobability methods for public opinion research, revisit past research in a new light, and present findings from our most recent experiment comparing alternative statistical estimation procedures across sample providers and survey topics.

13:45–14:15

A Partially Successful Attempt to Integrate a Web-Recruited Cohort into an Address-Based Sample

Phillip S. Kott (presenter), Matthew Farrelly, Kian Kamyab

(RTI International)

pkott@rti.org

A web-and-mail survey was conducted in Oregon on attitudes towards and use of recently-legalized marijuana. Roughly two-thirds of the respondent sample was selected via a

simple random sample of addresses. Sampled individuals were encouraged to respond by web, but about half of the respondents returned a mail questionnaire instead. Another third of the respondent sample was nonprobability, recruited via Facebook and responding by web. Thus, there were three cohorts: a mail cohort, a mail-to-web cohort, and a recruit cohort. Preliminary investigations revealed that the recruit cohort did not look like the mail cohort, but that the recruit cohort might be similar to the mail-to-web cohort. The paper demonstrates how and why the SUDAAN procedure WTADJX was used to calibrate the randomly-selected respondents to variable totals from the American Community Survey while the mail-to-web and recruit cohorts were calibrated to each other using the ACS variables and political affiliation. WTADJX was used to assess whether differences between estimates from the mail-to-web and recruit cohorts were statistically significant. The calibrated weights for these cohorts were then scaled so that the population they represented was single-counted. Finally, delete-a-group jack-knife weights were developed for estimates computed from the entire respondent sample.

14:15–14:45

A test of sample matching using a pseudo-web sample

Jack Gambino, Golshid Chatrchi

(Household Survey Method Division, Statistics Canada)

jack.gambino@canada.ca, golshid.chatrchi@canada.ca

With increasing levels of nonresponse in household surveys, there is renewed interest in alternatives to the traditional way of conducting such surveys. Rivers (2007) proposed the sample matching approach, and showed under certain assumptions matching from a sufficiently large and diverse web panel provides results similar to a simple random sample. In this paper, we test the sample matching approach due to Rivers using a pseudo-web sample. We use data from two different household surveys to simulate the sample matching methodology. The population of the study consists of the 2011 National Household Survey (NHS) respondents and the Canadian Labour Force Survey (LFS) respondents are treated as a pseudo-web sample. Different matching techniques and variables are tested, and the robustness of the method is evaluated under various conditions. We also briefly describe an experiment that uses a real web sample to collect data for sample matching.

15:00–15:30

Expanding the toolbox: inference from non-probability samples using machine learning

Joep Burger, Bart Buelens, Jan van den Brakel

(Department of Methodology, Statistics Netherlands, Heerlen, the Netherlands; Department of Quantitative Economics, Maastricht University, Maastricht, the Netherlands)

j.burger@cbs.nl, b.buelens@cbs.nl, ja.vandenbrakel@cbs.nl

Social and economic scientists are currently exploring non-probability samples like big

data as an alternative to traditional survey samples. Big data generally cover an unknown part of the population of interest. Simply ignoring this potential selection bias is error-prone. The mere volume of data provides no guarantee for valid inference. Tackling this problem with methods originally developed for probability sampling is possible but shown here to be limited, since they often fail to account for the data generating process. We propose a more general predictive inference framework, including three classes of inference methods: design-based, model-based and machine learning techniques. The machine learning methods we studied are k-nearest neighbor, artificial neural networks, regression trees and support vector machines. In a simulation study, we create selective samples from real-world data on annual mileages by vehicles, infer a population parameter using these inference methods, and compare the method performances. Our results show that machine learning methods can outperform the other methods in removing selection bias. Describing economies and societies using sensor data, internet data, social media and voluntary opt-in panels can be cost effective and timely compared with traditional sample surveys, but require inference procedures that account for the data generating process.

15:30–16:00

Investigation into the use of weighting adjustments for non-probability online panel samples

Dina Neiger, Darren W. Pennay, Andrew C. Ward, Paul J. Lavrakas

(ANU Centre for Social Research and Methods, Australian National University; Institute for Social Science Research, University of Queensland; NORC at the University of Chicago; Office of Survey Research at Michigan State University)

dina.neiger@srcentre.com.au, darren.pennay@srcentre.com.au,

andrew.ward@srcentre.com.au, pjlavrakas@centurylink.net

Weighting is used to try to reduce total survey error for probability samples by making adjustments for selection probability and enforcing population distribution across key demographics.

There is no agreement on the efficacy of similar weighting adjustments for correcting bias of non-probability samples given non-probability selection methods, enforcement of quotas and the proprietary mechanisms used by sample providers to ensure that their sample resembles the population.

Alternative methods, such as blending and calibration (e.g. DiSogra et al. 2011) and propensity-based weighting (e.g. Schonlau et al. 2003) have shown benefit but there is limited research available comparing the impact of different methods on the total survey error.

Our presentation aims to contribute to this topic through a comparative evaluation of weighting alternatives by using data from the recent Australian Online Panels Benchmarking study (Pennay et al. 2016). Survey items included in the study were selected to allow comparison with many demographic, health and wellbeing benchmarks. The availability of these official benchmarks makes it possible to evaluate a range of methods

with respect to their impact on the total survey error.

The presentation will summarise the results of our evaluation and discuss alternatives methods for weighting adjustments to nonprobability samples.

16:00–16:30

A bootstrap method for estimating the sampling variation in point estimates from quota samples

Jouni Kuha, Patrick Sturgis

(London School of Economics and Political Science; ESRC National Centre for Research Methods, School of Social Sciences, University of Southampton)

P.Sturgis@soton.ac.uk, j.kuha@lse.ac.uk

Measures of uncertainty in survey estimates which are derived under assumptions of probability sampling are not directly applicable to quota samples, yet ignoring the sampling variability in quota sample estimates is also clearly unsatisfactory. We propose a method of calculating the precision of estimates from quota samples which better reflects their sample design and conveniently accommodates the features of the estimation applied to the samples. This is a bootstrap re-sampling method which involves the following steps: (i) draw independent samples by sampling respondents from the full achieved sample, in a way which mimics the quota sampling design; (ii) for each sample thus drawn, calculate the point estimates of interest in the same way as for the original sample; and (iii) use the distribution of the estimates from the samples to quantify the uncertainty in the survey estimates. We illustrate the method and assess its performance relative to existing approaches by application to opinion poll estimates of vote shares prior to the 2015 UK General Election.

16:45–17:45

Looking for rigor in all the wrong places

Andrew Gelman

(Columbia University, New York, USA)

gelman@stat.columbia.edu

What do the following ideas and practices have in common: unbiased estimation, statistical significance, insistence on random sampling, and avoidance of prior information? All have been embraced as ways of enforcing rigor but all have backfired and led to sloppy analyses and erroneous inferences. We discuss these problems and some potential solutions in the context of problems in applied survey research, and we consider ways in which future statistical theory can be better aligned with practice.

Day 2, March 17th 2017

9:00–9:30

The Accuracy of Online Surveys: Evidence from Germany

Annelies G. Blom, Daniela Ackermann-Piek, Susanne Helmschrott, Carina Cornesse, Christian Bruch, Joseph W. Sakshaug

(Department of Political Science, School of Social Sciences, University of Mannheim; Collaborative Research Center 884 'Political Economy of Reforms', University of Mannheim; GESIS – Leibniz Institute for the Social Sciences; University of Manchester, Manchester, UK)

blom@uni-mannheim.de, daniela.ackermann@uni-mannheim.de,
helmschrott@uni-mannheim.de, carina.cornesse@uni-mannheim.de,
christian.bruch@uni-mannheim.de, joesaks@umich.edu

Online surveys have become more and more important during the past years. They promise a faster and cheaper data collection, enable researchers to react to societal events within days, and, due to their self-completion format there are no interviewer effects and social desirability biases can be reduced. However, despite the ubiquity of the internet and emails in our daily lives, we still cannot sample individuals or households directly online, because no frames of email addresses or internet access points are available. For probability online surveys, we thus have to sample via initial probability face-to-face or telephone interviews, which is costly. This lack of available sampling frames paired with the attractiveness of the online mode has given rise to an industry of nonprobability online surveys.

This study assesses the accuracy of eight nonprobability online samples with the accuracy of two probability online samples, and compares these to two gold-standard probability face-to-face samples in Germany. All samples were specifically drawn to be representative of the general population aged 18 to 70 in Germany. We compare aggregate results against official benchmarks on socio-demographic characteristics and political participation. The probability samples showed higher accuracy than nonprobability samples. Additional weighting reduced differences between the samples.

9:30–10:00

Assessing the accuracy of 51 non-probability online panels and river samples: A re-analysis of the Advertising Research Foundation (ARF) online panel comparison experiment.

Mario Callegaro, Yongwei Yang, Katherine Chin, Ana Villar, Jon A. Krosnick

(Brand Studio, Research at Google UK)

callegaro@google.com

Survey research is increasingly conducted using online panels and river samples. With a large number of data suppliers available, data purchasers need to understand the accuracy of the data being provided and whether probability sampling continues to yield more accurate measurements of populations. This paper evaluates the accuracy of a probability sample and non-probability survey samples that were created using various different quota sampling strategies and sample sources (panel versus river samples)

on the accuracy of estimates. Data collection was organized by the Advertising Research Foundation (ARF) in 2013. We compare estimates from 45 U.S. online panels of non-probability samples, 6 river samples, and one RDD telephone sample to high-quality benchmarks – population estimates obtained from large-scale face-to-face surveys of probability samples with extremely high response rates (e.g., ACS, NHIS, and NHANES). The non-probability samples were supplied by 17 major U.S. providers. The online samples were created using three quota methods: (A) age and gender within regions; (B) Method A plus race/ethnicity; and (C) Method B plus education. Comparisons are made using unweighted and weighted data, with different weighting strategies of increasing complexity. Accuracy is evaluated using the absolute average error method. The study illustrates the need for methodological rigor when evaluating the performance of survey samples.

10:00–11:00

The perils of non-probability sampling

Jelke Bethlehem

(Leiden University, Institute of Political Science)

`bethlehem@xs4all.nl`

Ever since the 1940s the guidelines of good survey research strongly advise to apply random sampling, as this makes it possible to generalize from the sample to the population. If the principles of probability sampling have been applied, it is always possible to compute valid estimates of population characteristics. Moreover, the accuracy of the estimates can be computed by means of confidence intervals or margins of error.

Developments in society face the survey researcher with new challenges. One of the problems is the increasing nonresponse rates, which affect the validity of surveys. Another problem are surveys costs. High quality surveys (for example CAPI surveys) are very expensive. Therefore, researchers are looking for cheaper alternatives. Also, for some surveys (for example CATI surveys) it is hard to find proper sampling frames.

And then came the internet. It made it possible to conduct online surveys. The advantages of online data collection (it is fast, simple, and cheap) on the one hand, and the lack of proper sampling frames on the other, caused many online surveys to be self-selection surveys. This is a form of non-probability sampling. Self-selection surveys have disadvantages. Estimates may be invalid, and it is impossible to compute the accuracy of estimates.

The presentation compares surveys based on random sampling with those based on self-selection. It also attempts to answer the question whether a probability sample with a substantial amount of nonresponse is not just as bad as a non-probability sample based on self-selection. Some examples show the perils of this type of non-probability sampling.

11:15–11:45

An Empirical Process for Using Non-probability Survey for Inference

Robert Tortora, Ronaldo Iachan
(ICF)

Robert.Tortora@icf.com, Ronaldo.Iachan@icf.com

While non-probability sampling (NPS) surveys are widely in use in market research, their adoption for official statistics is much more problematic. The adoption and acceptability of NPS surveys seem linked to assurances of the quality (or accuracy) of the NPS data. To date most of the research involves comparisons to probability survey estimates or uses some form of modeling derived from a probability survey to produce estimates. There had been little research on going beyond the comparison stage where a NPS stands alone and is valid for statistical inference. This paper describes a two-step empirical method that first compares an NPS survey, or series of survey, from an online panel to a probability survey. The second step proposes how, at a later date, the NPS survey can stand alone for statistical inference. The approach also relies on defining a priori rules allowing the data user to decide on the level of risk they are willing to accept for a satisfactory comparison at the first step. We use two different online samples for a large urban area: a traditional quota sample and a sample based on filling the most problematic quotas first. Here no follow-up emails are sent and new invitations are sent until all the quotas are filled. The key aspects of the methodology include transparency through an a priori decision rule motivated by the ASPIRE system developed by Bergdahl et al. (2014). For the first step we propose creating a scoring index based on 1) overall survey estimates, 2) subgroup estimates and 3) the ratio of coefficients of variation of the post-stratification weights from the NPS and the probability survey. A predetermined cutoff value determines the risk accepting or rejecting the NPS estimates. Assuming a successful comparison at step 1 we again define an a priori rule that compares the demographics of the online panels' target population to the demographics at a later time for the stand alone conduct of the NPS survey using the same methods. We illustrate our step 1 empirical method by comparing data from the two NPS samples to a probability survey of the same area.

11:45–12:15

Inbound Call Survey (ICS) – A New Methodology

Karol Krotki, Burton Levine, Georgiy Bobashev, Scott Richards
(RTI; Reconnect Research)

kkrotki@rti.org

Inbound call methodology is based on the possibility of intercepting incorrectly-dialed calls and replacing the curt termination message with an invitation to complete a survey. The methodology is nonprobabilistic and open to bias but on the other hand it is extremely inexpensive and quick. The number of such calls that can be intercepted on a daily basis in the USA and Canada numbers in the millions. Callers hear an intercept message such as “Please take our national health survey. Your call couldn’t be completed

and was redirected to this survey”. Multiple modes can be used for inbound call surveys including IVR (Interactive Voice Response), a live interviewer, or redirection to a web site to complete a web based instrument.

We first outline the methodology and how we weight adjust the ICS data to known population totals using calibration. Next, we report on the methodology to compare ICS results with established national surveys (2015 American Community Survey, 2015 National Health Interview Survey) and the bias that was found. We quantify bias by treating the population estimates from the ACS and the NHIS as correct and quantify the deviation from the unweighted ICS results for demographic characteristics and the weighted ICS results for the health outcomes. We also examine bias in a multivariate analysis. We show how ICS methodology can produce estimates with mean squared error comparable to an outbound telephone survey. Furthermore, we show how these gains in mean squared error are achieved at considerably lower cost. We also discuss the ICS as an efficient means of screening for rare and hard-to-reach population as well as a tool for bio-surveillance.

We show that while the results of the comparisons are promising, more rigorous research is needed to address potential biases, some of which are related to the timing of survey and the way questions are asked. We also discuss issues related to questionnaire design, sensitive topics, informed consent, and the protection of human subjects.

12:15–12:45

In search of best practices

Sander Steijn, Joost Kappelhof

(The Netherlands Institute for Social Research/SCP)

s.steijn@SCP.NL, j.kappelhof@SCP.NL

The Netherlands Institute for Social Research/SCP conducts sociocultural research in the Netherlands. This type of research, frequently based on surveys, covers a wide range of topics, such as health, education, sociocultural integration, discrimination, etc., among a variety of groups living within the Netherlands. Quite often the research is targeted at difficult to survey groups such as the elderly, children, ethnic minorities or sexual minorities. These groups can be difficult to survey. Among the elderly, health or cognition can provide challenges, surveying people in institutions is hindered by legal and coverage issues and surveying ethnic minorities can entail cultural difficulties. These issues, in turn, can lead to increased coverage, nonresponse or measurement error.

A different type of problem the SCP faces, concerns so called ‘hidden’ populations such as the LHBT community, where the lack of a useable sample frame can make it nearly impossible to draw a probability based sample. All of the aforementioned problems can be further complicated by (internal and external) demands for timely reports on findings. As the generalizability of results is very important for the SCP, non-probability based samples have, in the past decades, not been a typical choice for SCP-research. However, in the light of the concerns that were addressed in the previous paragraphs, in the recent years the SCP has in some cases opted for non-probability based samples. In these

instances, a balance needed to be struck on addressing the most urgent and relevant research questions and the wish to make generalizations on research findings. In our presentation we will briefly describe a few examples of studies that made use of a non-probability based sample and discuss how we dealt with the issues of generalizability of these results. On the basis of these examples, we formulate several ‘best practices’ on how to best deal with expectation management when presenting survey results from nonprobability samples.

Information

Venue

SciencesPo
254 Boulevard Saint-Germain
75007 Paris
France
Conference room: *Salle du Liepp* (first floor)

Registration

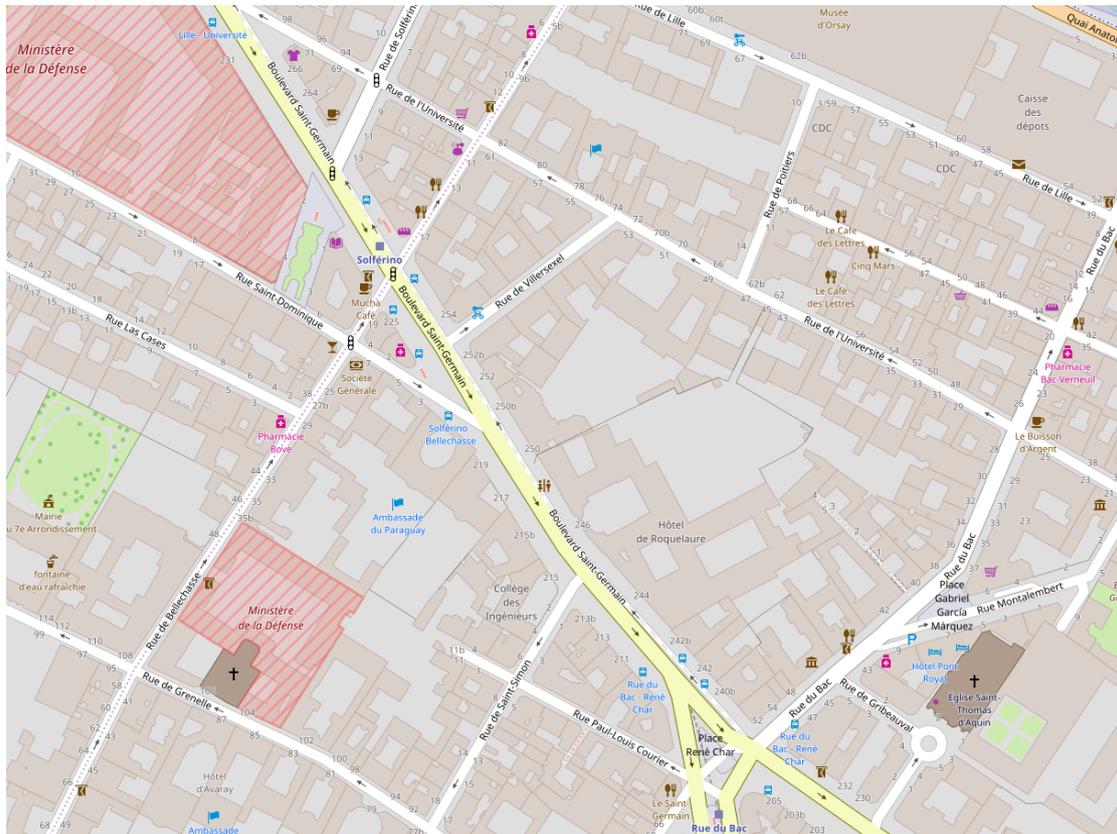
For Registration, please create an account on the ESRA-website at

<http://www.europeansurveyresearch.org>

Once registred, please log in at the ESRA website and select the Non Probability Conference. Click through and indicate whether or not you are coming to the conference dinner. Finally click pay for the conference fee.

The registration fee of € 50 is meant to cover costs for coffee, refrechments and to pay the persons at the registration desks. Costs for the conference dinner on the evening of March 16th 2017 are not included.

Map



Public Transport

Metro: Solférino (Ligne 12)

RER: Gare du musée d'Orsay (Ligne C)

Bus: Solférino (Ligne 63, 68, 69, 73, 83, 84, 94)

Bus, tram and subway tickets are cheaper if you buy them by 10 (“carnet de 10 tickets”, €16) There are also day passes (“Paris visite pass”), the one for two days cost €18.95; see http://www.ratp.fr/en/ratp_r_61584/tickets/.

Conference dinner

There will be a conference dinner on Thursday evening. Please indicate with your registration whether you like to participate in the dinner.

Wifi

Wifi will be provided at the conference venue.