PLATTFORM ftEval
Forschungs- und Technologieevaluierung

Nr. 37

Dec 2011

BUNDESMINISTERIUM FÜR WISSENSCHAFT UND FORSCHUNG
Minoritenplatz 5, A-1014 Wien
Mag. Irene Danler irene.danler@bmwf.gv.at
Mag. Simone Mesner simone.mesner@bmwf.gv.at

FWF – DER WISSENSCHAFTSFONDS
Sensengasse 1, A-1090 Wien
Dr. Rudolf Novak
novak@fwf.ac.at

BUNDESMINISTERIUM FÜR VERKEHR, INNOVATION UND TECHNOLOGIE
Renngasse 5, A-1010 Wien
Dr. Rupert Pichler rupert.pichler@bmvit.gv.at
Mag. Mario Steyer mario.steyer@bmvit.gv.at

JOANNEUM RESEARCH FORSCHUNGS-GESELLSCHAFT MBH
Sensengasse 1, A-1090 Wien
Mag. Wolfgang Polt wolfgang.polt@joanneum.at
Mag. Michael Dinges michael.dinges@joanneum.at

BUNDESMINISTERIUM FÜR WIRTSCHAFT, FAMILIE UND JUGEND
Stubenring 1, A-1010 Wien
Mag. Sabine Pohoryles-Drexel
Sabine.Pohoryles-Drexel@bmwfj.gv.at

KMU FORSCHUNG AUSTRIA
Gusshausstraße 8, A-1040 Wien
Dr.in Sonja Sheikh
s.sheikh@kmuforschung.ac.at

AQA – ÖSTERREICHISCHE QUALITÄTSSICHERUNGSAGENTUR
Wickenburggasse 26, A-1080 Wien
Mag. Alexander Kohler alexander.kohler@aqa.ac.at

LUDWIG BOLTZMANN GESELLSCHAFT
Nußdorfer Str. 64, A-1090 Wien
Mag.a Claudia Lingner
c.lingner@ludwigboltzmanngesellschaft.at

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY GMBH
Donaucitystraße 1, A-1220 Wien
Mag. Petra Wagner-Luptacik petra.wagner@arcs.ac.at
Dr.in Susanne Giesecke susanne.giesecke@arcs.ac.at

RAT FÜR FORSCHUNG UND TECHNOLOGIEENTWICKLUNG
Pestalozzigasse 4/DG 1, A-1010 Wien
Dr. Johannes Gadner
j.gadner@rat-fte.at

AWS – AUSTRIA WIRTSCHAFTSSERVICE-GESELLSCHAFT MBH, Ungargasse 37, A-1030 Wien
Wolfram Anderle w.anderle@awsg.at
Mag. Norbert Knoll n.knoll@awsg.at

TECHNOPOLIS AUSTRIA GMBH
Rudolfplatz 12/11, A-1010 Wien
Mag.a Katharina Warta
warta@technopolis-group.at

CDG – CHRISTIAN DOPPLER FORSCHUNGSGESELLSCHAFT
Sensengasse 1, A-1090 Wien
DI Mag.a Brigitte Müller mueller@cdg.ac.at

WIFO – ÖSTERREICHISCHES INSTITUT FÜR WIRTSCHAFTSFORSCHUNG, Arsenal, Objekt 20, A-1103 Wien
Dr. Andreas Reinstaller andreas.reinstaller@wifo.ac.at
Dr. Jürgen Janger juergen.janger@wifo.ac.at

CONVELOP KOOPERATIVE KNOWLEDGE DESIGN GMBH
Bürgergasse 8-10/I, A-8010 Graz
Mag. Markus Gruber
markus.gruber@convelop.at

WWTF – WIENER WISSENSCHAFTS-FORSCHUNGS– UND TECHNOLOGIEFONDS
Schlickgasse 3/12, A-1090 Wien
Dr. Michael Stampfer michael.stampfer@wwtf.at
Mag. Klaus Zinöcker klaus.zinoecker@wwtf.at

FFG – ÖSTERREICHISCHE FORSCHUNGS-FÖRDERUNGSGESELLSCHAFT, Sensengasse 1, A-1090 Wien
Mag. Klaus Schnitzer klaus.schnitzer@ffg.at
Mag. Leonhard Jörg leonhard.joerg@ffg.at

ZIT – ZENTRUM FÜR INNOVATION UND TECHNOLOGIE GMBH
Eberndorferstraße 4/DG 1, A-1010 Wien
Robert Mayer-Unterholzner robert.mayer@zit.co.at

ZSI – ZENTRUM FÜR SOZIALE INNOVATION
Linke Wienzeile 246, A-1150 Wien
Dr. Klaus Schuch schuch@zsi.at

Der Newsletter der Plattform Forschungs– und Technologieevaluierung ist ein unregelmäßig erscheinendes offenes Forum zur Diskussion methodischer und inhaltlicher Evaluierungsfragen in der Forschungs- und Technologiepolitik. **© Wien 2011, ZVR-Zahl: 937261837, ISSN-Nr.: 1726-6629**

Rupert Pichler

Preface

Once again this newsletter presents a broad range of topics. The discussion of recent evaluation results is complemented by papers discussing the impact of evaluations themselves. The impact of evaluations on policy makers' decision-making may often by hampered by unrealistic expectations. Karin Grasenick and Stephan Kupsa show in their paper that the complexity of the innovation system does not allow evaluations to yield simple answers, immediately feeding into the policy cycle. However, the usefulness of evaluation results also depends on the questions asked. In their analysis of possible peer-review related distortions within the Austrian Science Fund's project selection procedures Christian Fischer and Falk Reckling present some rather to-the-point findings. The very problem of peer-review biases is then dealt with by Christian Reiner who – focussing on the level of programme evaluations – proposes an ambitious randomized approach in order to control for those biases. The potential usefulness of evaluation results depends, of course, also on the methods employed. Sonja Sheikh, Sabine Mayer and Peter Kaufmann compare different methods of impact measurement with a closer look at matching approaches.

The following papers present recent programme evaluations which also aimed at delivering relevant information for decision-making. The two evaluation studies performed by Technopolis made significant use of qualitative contextualizing methods. The evaluation of the innovation voucher scheme *(Barbara Good, Brigitte Tiefenthaler)* shows that the increase in the number of companies going for more ambitious funding schemes after the voucher was not as high as expected. Nonetheless, the innovation voucher apparently became a useful scheme in its own right. The conclusions from the evaluation of the R&D headquarter programme *(Anton Geyer, Brigitte Tiefenthaler)* turned out to be more controversial. This is owed not only to the results themselves but also to the rather small number of cases which made the analysis of cause-impact relationships easier to understand – and, as a consequence, to criticize – than econometric approaches. Finally, Ilse Marschalek, Katharina Handler and Katharina Strasser analyse the attitude of young people towards nano technologies in the EU framework programme NANOYOU project, benefitting from a mix of various methods.

Ruper Pichler
Federal Ministry for Transport, Innovation and Technology
MAIL rupert.pichler@bmvit.gv.at

Karin Grasenick and Stephan Kupsa

Everything you always wanted to know about evaluation, but were never told

1

Introduction economic and social policy objectives

Our society uses political parties to establish which objectives are capable of winning a majority, which values underlie these objectives, and which responsibilities should be controlled by the state. To this end, visions are identified at the political level – long-term goals, models for a 'better society'. To actually implement them, it is essential for politics and administration to cooperate and interact. Accordingly, ministerial departments list long term goals in their external presentation:

- 'We want to strengthen Austria's economic competitiveness' *(cf. Austrian Federal Government, 2011)*;
- 'The goal of our economic policy is to constantly improve the population's quality of life (...)' *(cf. Federal Ministry of Economy, Family and Youth)*;
- 'Living longer in good health is our goal.' *(cf. Federal Ministry of Health)*;
- 'Future generations are to grow up in a society that offers equal opportunities for men and women.' *(cf. Federal Ministry for Transport, Innovation and Technology)*.

Democratic processes and structures, combined with a multiplicity of *(social)* policy issues, produce a complex system involving players from a large range of subsystems for political interventions *(cf. Grasenick, Wagner and Zumbusch, 2008)*. Moreover, many issues cut across multiple sectors that may be assigned to multiple ministerial remits, thus intensifying the complexity of the overall system. These different economic and social policy issues show the same commonalities in terms of complex coordination and control, i.e. their governance structures. Willke supplies a good definition of governance: *'Governance is the activity of coordinating communications in order to achieve collective goals through collaboration.' (Willke, 2007)*.

The road from identifying visions to taking concrete steps to implement them is full of obstacles. Considering the variety of players involved and the complexity of the issues, we need first of all to consider two aspects:

- It is important to understand how issues and 'concepts' are defined by the players and which factors of influence are relevant to them: How do players understand 'economic prosperity' or 'economic competitiveness'? What is meant by 'equal opportunities'? What contributes to competitiveness or equal opportunities?
- It is important to rely on basic theories, hypotheses or models on impact chains. Such models provide a differentiated idea of conceivable approaches, potential levers and relevant target groups. They strive to depict the interaction between social framework conditions *(metalevel such as laws and traditions)*, relevant subsystems *(mesolevel such as NGOs and enterprises)* and individuals *(microlevel)*. This interaction between levels and their factors of influence is, in the final analysis, decisive for the overall outcome.

Drawing on such underlying considerations, decision-makers in politics and administration ultimately need to decide on those measures which they expect to have the greatest leverage effect in order to achieve the goals for a given target group. Generally – and regardless of how issues are assigned to subjects – we can start out from the three levels noted above:

- at the metalevel, overriding specific programme targets: the legal framework *(e.g. occupational health and safety regulations, anti-discrimination laws, taxes, etc.)*;
- at the mesolevel: incentives for subsystems such as networks, organisations, associations, etc. *(support of health and safety measures at enterprise level, support of programmes for the structural implementation of equal opportunities, etc.)*;
- at the microlevel: individual support *(preventive health care for individuals, advice to founders, qualification programmes and grants, etc.)*.

Programmes to support the implementation of politically relevant objectives

Harmonisation processes between players and subsystems result in a mixture of interventions aimed to help achieve the metagoals at the levels and spheres of responsibility identified. Part and parcel of such interventions *(and a tool very widely used in Austria)* are funding programmes – and this appears to apply not just to research, technology and innovation policy *(cf. WIFO, KMU Forschung Austria, prognos and convelop, 2009)*, but as much to other policy fields such as economic, health and labour market policies *(cf. WIFO and Vienna University of Economics and Business, 2010)*.

Due to the interrelation of the overall factors of influence, individual programmes by themselves can only foster specific aspects and provide specific incentives. Nevertheless it is legitimate to define programmes suitable for a specific sphere of responsibility and make their contribution visible. Consequently there is a tendency to add metagoals as well as impact targets when designing programmes *(e.g. regional economic competitiveness, equal opportunities of women and men)*.

On the difficulty of assessing the effect of programmes

At some point during the course of a programme somebody will ask about the quantitative contribution made by the programme to achieve the political visions. The less experienced the players are in evaluating processes, the quicker this question will pop up.

[1] The contribution is the result of ongoing experience obtained from evaluating programmes in a range of different policy fields and discussions with representatives from administration on the design of visions, goals, targets, funding strategies and their evaluation.

The point of an evaluation is to answer such questions and to supply concrete proposals to improve performance – possibly as a help to decide on whether to continue or terminate a given programme. To arrive at a decision it is usually necessary to obtain quantitative findings of the effects as soon as possible. The cost of an evaluation needs to be proportionate to the funds used by the programme. For an evaluation to comply with expectations, we first need to face other issues than those of measurable impacts for a political vision, i.e.:

- Which goals were named in the programme; what is measurable and can be allocated to the programme?
- Which time frame can be monitored?
- Can control groups be taken into consideration? How can we assure that a comparison *(considering the many factors of influence)* is actually to the point?
- What is feasible and useful in the given financial framework?

The following example illustrates the dilemma faced by evaluations:

Has there been a change in the proportion of women and men in leadership positions?

The goal of raising the share of women in leadership positions helps towards achieving equal opportunities. Career development is affected by a large number of factors. At the microlevel these are biographical factors such as training experience which impact on the individual's understanding of him/herself, on the growth of self-confidence and on one's own role model. At the mesolevel there are cultural factors, such as the prevalence of existing role models in organisations and among superiors and their structural effects. And on the macrolevel there are social framework factors such as the distribution of family tasks, child care services and control mechanisms such as quota systems *(cf. for the scientific field: Shalala et al., 2006 and, frequently quoted in studies, Deutscher Wissenschaftsrat 2007, Stutz, H., Dubach, P., Guggisberg, J., Fuchs, G. and Strub. S., 2008).*

Programmes to promote equal opportunities thus include, as a rule, many approaches – they may start out at the microlevel, e.g. by enabling women to make use of specific grants, mentoring or coaching services that aim to provide them with the prerequisites to climb the career ladder.

Even though they are basically aimed at the individual, such programmes, when successful, can impact on the mesolevel as well. Experience thus gained may raise the willingness of mentors and bosses to increase the number of women employees in the future, or stimulate structural changes within an organisation to promote men and women at equal terms. Such effects are conceivable as well as desirable – but cannot be controlled by programmes aimed at the individual. Direct effects on the macrolevel are not expected, considering the multiplicity of factors and 'smallness' of the measures involved.

As a direct effect, a measure can contribute to upskilling or assist strategic career planning (in the case of mentoring and coaching). Whether women actually advance their career during the monitoring period, thus increasing the number of women in leadership positions, however depends on other factors *(i.a. availability of suitable jobs, number of qualified applicants for such jobs, etc.)*.

The desired contribution to an individual's career advancement thus is not necessarily observed during the monitoring period. And even where it can be observed there is still the question of how much a specific programme contributed to such change.

Nothing is also known about which course a career would have taken without the programme. Generally it would be possible to use comparison groups to study this issue. To this end it would be necessary to find and compare suitable matching pairs, i.e. women in a similar initial situation *(family, personal motivation, education/training and experience, current and desired position, organisational environment, etc.)* and study the development of their respective careers with and without the use of the programme to be evaluated. Such a method requires considerable input *(cf. Pointner, 2001)* and is limited in its informative value.

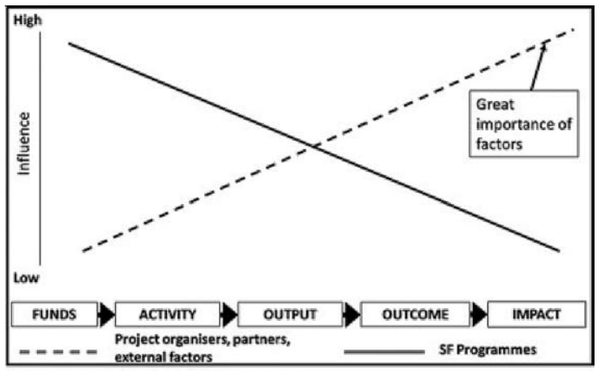
The challenge: evaluating small contributions towards major goals in a complex world

From our evaluation practice we therefore perceive several challenges in identifying and documenting the effect of programmes or measures:

- **Target levels:** Evaluations are frequently faced with the dilemma that programmes are formulated as contributions towards achieving goals or (meta)targets which are not within their direct sphere of influence.
- **Multiple influencing factors:** Programmes focus on specific targets. These targets are usually affected by a large range of factors. Social processes are typically complex and equivocal. Implementing a programme is just one of many factors/contributions to achieve the targets. These factors interact within the system, reinforcing but also cancelling out each other. Ultimately, however, it is the interaction of all factors as an overall system which determines that a given effect is achieved. Thus it is conceivable that a programme, while showing positive effects, still has these effects levelled by contrary factors. Yet this interaction between factors cannot be controlled by the measures and programmes to be evaluated.
- **Smallness of programmes:** This is closely linked to the fact that many programmes, while limited in their resources, still aim for a large scope in terms of reach and impact. Programmes thus may show quite positive results *(output, outcome)* within their limited sphere of action but these effects do not produce any quantitative changes in the metagoals.
- **Methodological limits:** The problems are of a financial nature (the financial framework does not allow for an in-depth analysis of all effect levels). Yet even where financing is sufficient, evaluators meet with great methodologi cal difficulties as the measures and target groups studied mutually influence each other.
- **Time horizon:** Measuring the impact is further complicated by the time lag. Change needs time. Sustainable change can thus be expected only after some years. This, however, makes it difficult to visualise a direct link between cause and effect.
- **Complex cause-effect relationship:** Even well-considered impact chains are based on plausibility considerations. Feedbacks and interdependencies can be identified only to a limited extent *(i.a. due to inadequate data)*.

In Figure 1, Earl, Carden and Smutylo provide a schematic diagram *(to be applied to structural funds)* showing that the effect of external factors increases from the provision of funding to the impacts, while the influence of the programme itself declines.

Figure 1
Importance of influencing factors.



Source: Earl, Carden and Smutylo, 2001, revised by convelop

Montague and Young differentiate between three 'circles of influence' *(cf. Montague, Young and Montague, 2003)*:

- Direct control – where the programme has fairly direct control of the results, typically the output level,
- direct influence – where the programme has a direct influence on the expected results, such as the reactions and behaviours of its clients through direct contact, typically the immediate outcomes,
- indirect influence – where the programme can exert significantly less influence on the expected results due to its lack of direct contact with those involved and/or the significant influence of other factors.

Conclusion: have the courage for a realistic outlook

In order to counter the dilemma of issues posed and practical challenges with a positive spirit and thereby achieve an improvement in evaluation results and in the effect of the evaluation itself, we offer several approaches:

1. **Focus the programme goals on and measure them against an actually feasible sphere of impact.** The programmes and their targets should be concentrated and limited strictly to those fields that can be directly influenced. A concrete target could be to support specific cooperation projects between new partners in science and business. Metagoals outside the programme's sphere of influence should not be used as programme targets. Rather, you can formulate an expectation, based on the underlying 'hypothesis', that a *(non-quantifiable)* contribution is made towards achieving this metagoal. Programmes by themselves may produce positive impact even if not visible at the 'macrolevel' because they are overlaid by other influencing factors.
2. **Concentrate programme evaluations on what is practicable.** Outputs can be directly influenced and directly ascertained by the programme's players. Outcomes *(i.e. indirect effects produced by the use of direct programme outputs, frequently aiming at behaviour change)* can be analysed only after a time lag and, moreover, by considering the many influencing factors. Where quantitative impacts are to be mapped it is necessary to refer to a longer time period.

3. **Refocus on overriding systemic evaluation.** In order to cope with the complexity described above we suggest thinking more in terms of comprehensive 'intervention spaces' *(where the development of metagoals and targets is considered in its totality within a long-term reference system)* already when analysing outcomes and even more when analysing impacts. The results of such systemic evaluation approaches should be more interesting, even for pecific programmes and measures, than any attempt to deal with issues regarding the overall system from the perspective of a given programme.

References

Bundesministerium für Gesundheit (Austrian Ministry of Health)
www.gesundheitsziele-oesterreich.at *(last viewed on 25 November 2011)*
Bundesministerium für Verkehr, Innovation und Technologie
(Austrian Federal Ministry for Transport, Innovation and Technology)
www.bmvit.gv.at/innovation/humanpotenzial/talente_nuetzen/index.html *(last viewed on 25 November 2011)*
Bundesministerium für Wirtschaft, Familie und Jugend
(Austrian Federal Ministry of Economy, Family and Youth)
www.bmwfj.gv.at/WIRTSCHAFTSPOLITIK/WIRTSCHAFTSPOLITIK/Seiten/default.aspx
(last viewed on 25 November 2011)
Carden F., Eearl S. and Smutilo T. (2001)
Outcome Mapping. International Development Research Centre, Ottawa
Deutscher Wissenschaftsrat (2007)
Empfehlungen zur Chancengleichheit von Wissenschaftlerinnen und Wissenschaftlern.
www.gleichstellung.uni-bonn.de/pdf-dokumente/empfehlungen_wissenschaftsrat
(last viewed on 28 November 2011)
Grasenick K., Wagner G. and Zumbusch K. (2008)
Trapped in a net: network analysis for network governance. VINE Vol. 38 No. 3
Mayne J. (2008)
Adressing Cause and Effect in Simple and Complex Settings Through Contribution Analysis. Discussion Draft.
www.rcep.ca/distribution/20080515_mayne_john_b.pdf *(last viewed on 28 November 2011)*
Montague S., Young G. and Montague C. (2003)
Using circles to tell the performance story. Canadian Government Executive 2: 12-16.
http://pmn.net/library/usingcirclestotelltheperformancestory.htm *(last viewed on 28 November 2011)*
Österreichische Bundesregierung (Austrian Federal Government) (2011)
Strategie der Bundesregierung für Forschung, Technologie und Innovation.
Pointner W. (2001)
Matching Pairs Analyse als mikroökonometrischer Evlaluierungsansatz zur Additionalitätsmessung. 11-16
Plattform Forschung Technologie Evaluierung (2001): Newsletter Nr. 13, Vienna
Shalala D., Agogino A., Baiyn L., Birgenau R., Cauce AM., Deangelis C., Denton D., Grosz B., Handelsman J., Koehane N., Malcom S., Richmond G., Rivlin A., Simmons R., Spelke E., Steitz J., Weyuker E. and Zuber M. (2006)
Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering.
The National Academies. www.nap.edu/catalog/11741.html *(last viewed on 28 November 2011)*

Stutz H., Dubach P., Guggisberg J., Fuchs G. and Strub. S. (2008)
Geschlecht und Forschungsförderung *(GEFO)*.
WIFO, KMU Forschung Austria, prognos u. convelop (2009)
Systemevaluierung des österreichischen Innovationssystems. Synthesebericht
WIFO and Vienna University of Economics and Business (2010)
Verwaltungsmodernisierung als Voraussetzung für nachhaltige Effizienzgewinne im öffentlichen Sektor
Willke H. (2007)
Smart Governance, Governing the Global Knowledge Society, Frankfurt/New York

Authors

Karin Grasenick
convelop cooperative knowledge design gmbh
MAIL karin.grasenick@convelop.at

Stephan Kupsa
convelop cooperative knowledge design gmbh
MAIL stephan.kupsa@convelop.at

Christian Fischer and Falk Reckling

The legitimacy of a funding agency in basic research

Some empirical evidences from the Austrian Science Fund (FWF)

The legitimacy of a funding agency in basic research depends at least on the followings factors:

1. **organisation's ability to minimise distortions in approval probability by its decision procedure**
2. **the scientific quality of results produced by funded research proposals**
3. **the acceptance of the procedures by the scientific community**

To ensure these requirements, the FWF started in 2010 a sequence of empirical studies which contain analyses of the decision making procedure *(peer review)* as well as statistical and bibliometric analyses of the outcome of FWF funded proposals. It will finally be round off by surveys among the Austrian scientific community in 2013. Some of the analyses are conducted by the FWF itself but additionally tested and supplemented by independent experts like a research group of the ETH Zurich and MPG Munich *(Hans-Dieter Daniel, Rüdiger Mutz and Lutz Bornmann)* or the bibliometric research group of the CWTS Leiden. All results are published as policy papers or in relevant international journals.

FWF' Decision making procedure

The first study produced in this context was published in 2010
Factors Influencing Approval Probability in FWF Decision-Making Procedures
FWF Stand-Alone Projects Programme, 1999 to 2008

The most important results are summarized as follows:

Motivation

The study was motivated by the insight that the legitimacy of decision-making procedures at funding agencies for basic research depends heavily on the organisation's ability to minimise distortions in approval probability – for example based on age, gender or research field – wherever possible. Due to human fallibility, erroneous decisions can never be ruled out entirely, but if systematic differences appear in the probability of funding approvals, then such differences either have to be eliminated by changing the decision-making procedure or they have to be made transparent and explainable. The purpose of this study was to examine any such systematic distortions in the FWF decision-making process.

[2] The study in full can be downloaded from FWF's homepage
www.fwf.ac.at/de/downloads/pdf/FWF-ApprovalProbability_P-99-08_15-12-2010.pdf

In order to select the most relevant among a large number of possible questions, we have collected a number of 'urban legends'. These refer to individual opinions and conjectures regarding the FWF's decision-making procedures as expressed by applicants, FWF decision-makers and employees as well as research policymakers in recent years.

These legends include statements such as the following:

- Certain disciplines, such as the humanities, are systematically placed at a disadvantage;
- Younger applicants have greater difficulties obtaining grants;
- Applicants who are not employed full time at an institution *(known as 'independent applicants')* are at a disadvantage compared to those who hold full-time positions at research institutions;
- Interdisciplinary proposals are less likely to succeed than monodisciplinary projects;
- 'Expensive' projects are less likely to succeed than 'inexpensive' ones;
- Female applicants are placed at a disadvantage. This can be attributed to distortions in areas such as the peer review process, the stronger interdisciplinary orientation of women in research, and/or the larger share of 'unstable' employment situations among women;
- Reviewers from countries a, b and c often assign lower ratings than their counterparts in countries x, y and z;
- There are more divergent reviews in the humanities and social sciences than in the natural sciences.

Such conjectures are to be taken seriously and used as a guide for the study.

Data

The analyses are confined to the Stand-Alone Projects Programme in the period from 1999 to 2008. Due to its size *(approximately 60% of all FWF grants)*, this programme can be considered representative of the FWF's decision-making procedures. Data were available of approximately 8,000 applications, which generated a total of 3,500 grants and 21,000 reviews. In order to depict development processes in specific cases, over 900 applications – resulting in 291 grants and 2,200 reviews – from the year 2009 were also included in the sample.

This study focuses on FWF applicants' probability of success based on the approval rate relative to the most important available characteristics of applicants *(including research field, age, gender, independent applicant status, funding amount requested)* and of reviewers *(location, gender, review rating)*.

Findings

The FWF examined its Stand-Alone Projects Programme over the period from 1999 to 2008 *(including a number of comparisons with the year 2009)* in order to determine approval probabilities based on various characteristics of the applicants and reviewers, thus testing the validity of its decision-making procedures.

The most important insights arising from this endeavour are presented in this discussion paper, which is intended to allow the scientific community, research policymakers and all other interested parties to discuss the findings and support the FWF in the continued optimisation of its procedures.

The key findings of the study are summarised below in the form of **questions (Q),** answers (A) and *problems (P)*.

Q: Are the humanities systematically placed at a disadvantage?
A: No. Along with mathematics, physics and biology, most humanities disciplines have the highest approval rates. Social sciences exhibit the lowest approval rates. The approval rates correspond roughly to the international performance of the respective scientific disciplines. Moreover, approval probability also hinges on each discipline's dependence on grants for basic research.
P: In order to strengthen the 'weaker' disciplines, it will be necessary (a) to allocate positions at research institutions according to transparent international performance criteria, (b) to make research funded by third parties more attractive by compensating research institutions for overhead costs, and (c) to provide greater relief for high-quality researchers from disciplines characterised by very high teaching workloads.

Q: Do younger applicants face greater difficulties than older applicants?
A: This used to be the case, but now it is no longer a problem. Until 2004, applicants under 35 years of age did exhibit a lower approval rate compared to older age groups. However, in recent years the FWF has successfully implemented a number of measures to balance out these differences.
P: However, it is still necessary to create career paths which enable junior scholars to gain independence in research quickly and to find permanent employment at research institutions after fulfilling transparent performance criteria. This is especially true in the case of female scholars.

Q: Are women placed at a disadvantage? If yes, why?
A: The share of female applicants has increased drastically in recent years, thus we can expect this share to increase to one-third or more in the medium term. However, women exhibited a lower average approval rate than men for stand-alone projects in the period from 1998 to 2008. This is most probably not linked to the decision-making procedure *(including the peer review process)*, but rather to the underrepresentation of women in certain disciplines, to additional *(especially family-related)* burdens during stages decisive to an academic career, and to the larger share of unstable employment relationships among women.
P: Measures to promote women in Austria should be reviewed to determine whether they provide successful applicants with support and resources (including research funding, child care and international mobility) which help compensate for competitive disadvantages at ages decisive to their careers. The FWF has already taken measures in this area, especially in its women's programmes.

Q: Are independent scientists (i.e. those who plan to finance their own salaries using FWF funds) at a disadvantage compared to researchers with permanent positions at research institutions?
A: No. In fact, independent applicants have slightly *(but not significantly)* higher chances of success in nearly all disciplines. However, the growing share of applicants in this category is problematic, especially if they migrate up through the age groups *(i.e. if scientists continue to rely on independent applications for excessively long periods and are unable to find permanent employment at research institutions)*. This is especially true in the case of female scholars.

P: It is helpful if junior scholars prove their performance potential early in their careers by successfully obtaining third-party funding for their own positions. However, this must not become a permanent arrangement. This means that research institutions should develop recruitment mechanisms which make successful third-party funding applications an important criterion in the allocation of positions.

Q: Do interdisciplinary applicants face greater difficulties than monodisciplinary applicants?
A: There is no simple answer to this question. It does appear that interdisciplinary applications face slightly greater obstacles compared to their monodisciplinary counterparts. However, this difference is only relevant in certain scientific disciplines, especially in cases where connections are established with 'weaker' disciplines.
P: It remains a challenge for funding agencies to determine (a) whether and (b) how to identify interdisciplinary projects in a more targeted manner, and (c) how decision-making procedures can then be adapted (specifically in order to prevent 'interdisciplinarity' from becoming a strategic tool in funding applications).

Q: How international is the FWF's review process? Are there any differences in review behaviour depending on the countries in which the reviewers work?
A: The FWF has continued to internationalise its review process in recent years and also makes efforts to base its reviewer choices on the scientific productivity of each country and region. In particular, the share of reviewers from German-speaking countries has been reduced significantly.
Reviewers from countries with high levels of scientific productivity tend to give more stringent reviews than reviewers from other countries.
In addition, the share of women among reviewers has increased, but it still does not match the share of female applicants in all research areas.
P: The FWF must continue to internationalise its review process. In the medium term, it will also be important to involve experts from emerging regions (including Asia and Latin America) more heavily in the review process. Finally, it will also be necessary to discuss how the share of female reviewers should develop in the coming years. This process will be faster in some disciplines (biosciences, humanities, social sciences and medicine) than in others (natural and technical sciences), in which women are still heavily underrepresented.

Q: Does the FWF receive more divergent ratings from reviewers in the humanities and social sciences than in the natural sciences, biosciences, technical sciences or medicine?
A: No. There are no significant differences between the research areas.
P: However, reviewers' assessments diverge at times in all research areas. In general, however, no funding agency has managed to resolve this problem up to now, and such disagreement is also an integral part of the scientific discussion process. In this context, funding agencies such as the FWF face the challenge of constantly reviewing and refining their procedures regarding the applicants' ability to respond/react to rejections (resubmission procedures).

Rüdiger Mutz, Lutz Bornmann, Hans-Dieter Daniel
Does Gender Matter in Grant Peer Review at the Austrian Science Fund (FWF)?
The data used for this study are independently analysed by Hans-Dieter Daniel, Rüdiger Mutz *(both ETH Zurich)*, and Lutz Bornmann *(Max-Planck-Society Munich)* with advanced statistical methods. They started their analyses with an investigation of the question whether gender matters in the FWF grant peer review and board of trustees` decision-making. The manuscript with the results of this investigation has been submitted for publication to a topical journal issue on 'Sex and Gender Differences Revisited'. They evaluated the external reviewers' ratings and board of trustees' final decision: approval or no approval for funding with respect to gender. In line with the current state of research, they found that the final decision was not significantly associated with applicant's gender or with any correspondence between gender of applicants and reviewers. However, the decisions on the grant applications showed a robust female reviewer salience effect. That means the final decision of the board of trustees varies with the proportion of female reviewers among all reviewers of a proposal *(salience)*. The approval probability decreases *(up to 10%)*, when there is a parity or majority of women in the group of reviewers.

Authors

Christian Fischer/Falk Reckling
Austrian Science Fund
MAIL falk.reckling@fwf.ac.at

Rüdiger Mutz/Hans-Dieter Daniel
ETH Zurich
MAIL mutz@gess.ethz.ch

Lutz Bornmann
Max Planck Society
MAIL lutz.bornmann@gv.mpg.de

Christian Reiner

Evaluating innovation policies by chance? The case for randomized R&D-programme evaluation

3

Introduction

The call for evidence-based policy making is a rational and perhaps the only possible scientifically justified answer to critics and advocates of state intervention. Expanding 'good' policies and contracting 'bad' ones can raise the efficiency of economic policies in a substantial way. Of course, 'good' and 'bad' are relative terms given the aims of the specific policy intervention. The remaining question is how to distinguish these two policies from each other? Yet, even if this question could be solved unambiguously, which is of course impossible even form a theoretical point of view, policy makers face several trade-offs between efficiency, equity and – above all – re-election. As a result, reality displays a mixture of 'good' and 'bad' policies. The aim of the social scientist is to bring facts to bear on these policies *(Jaffe 2002)*.

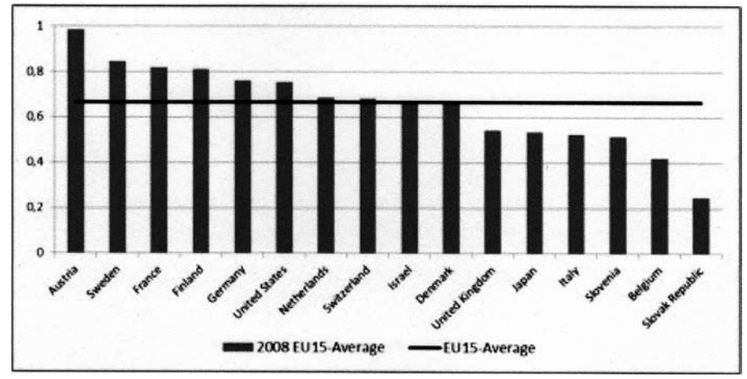
In this vein, the amount of public resources devoted to R&D which has been growing in the past decades in most OECD countries asks for thorough evaluations of their effectiveness *(Link and Scott 2011)*. Additionally, the limited growth prospective of the EU as well as the necessity to reduce public debts in the medium term render evaluation even more important about the effectiveness and efficiency of public spending in general *(including R&D spending)*. Yet, evaluations do not necessarily provide the rigorous results needed to inform policy makers on which of the tough choices they should make given, certain criteria of efficiency and equity. While there is a rather high frequency of evaluations in Austria, most of them remain on the level of descriptive and anecdotal evidence regarding the causal effect of public R&D programmes. Microeconometric evaluation is almost absent *(for rare exceptions see Polt and Pointner 2005, Streicher 2007)*. From a scientific point of view, this is a highly unsatisfactory situation. As a result, we do not know to what extent a positive correlation between a policy measure and the outcome is actually causal or simply the result of a selection bias, produced by multiple and complex selection mechanisms of R&D programmes.

Figure 1 shows that governments in the EU and OECD spend on an average 0.66% of GDP on gross expenditure on R&D *(GERD)*. The Austrian public sector displays one of the largest public expenditures in the field of innovation policy with roughly 1% of GDP.

[3] This article was written as part of the research project „Output-orientierte Evaluierung öffentlich geförderter FTI-Programme – Möglichkeiten und Grenzen', financed by the Austrian Federal Ministry for Transport, Innovation and Technology.

Figure 1
Government-financed GERD as a percentage of GDP 2008 or latest available

OECD, MSTI. Notes: The values from Sweden, Netherlands, Denmark and Belgium are from 2007.



Currently two avenues to improve this state of affairs are pursued: Firstly, better availability of survey data permits a range of non-experimental studies. The respective methodological literature has matured also in the field of R&D evaluation and there are now widely excepted and well-understood techniques available *(Czartnitzki et al. 2003; Cerulli 2010)*.

Secondly, experimental studies have been suggested in the literature *(Angrist and Pischke 2010)* as another possibility. While the first approach is clearly the more realistic one *(but still has some limitations)*, this paper investigates the arguments for and against 'social experiments' in the evaluation of public R&D programmes. Generally, experiments are among the most credible ways to detect causal effects. There is a considerable divide in the application of social experiments between Europe and the US. While Burgess *(1995:64)* states that' *(…)* classical experimentation on a modest scale has become an accepted part of policy evaluation in the United States', the opposite holds for Europe. European politicians are generally still reluctant to evaluate the causal effects of policies via social experiments. Nevertheless, work in the social sciences has delivered several reasons in favour of social experiments *(Schmidt 2007)*. For example, the Dutch economists Cornet and Webb ink *(2004)* argue that 'controlled experiments can be applied in many areas. They can yield advantages especially in recurrent policy issues, i.e. those policy issues where the effects of policy instruments have been debated for a long time.' Notably, as an example of these 'recurrent policy issues', they also mention R&D subsidies.

Yet, discussion and application of social experiments in the context of innovation policy is almost insignificant. Exemptions are e.g. Jaffe *(2002)*, Cornet et al. *(2006)* and Brezis *(2007)*. This result is also reflected by the fact that even comprehensive methodological compilations such as the EU's RTD Evaluation Toolbox *(2002)* do not devote specific attention to social experiments. Hence, this paper aims to support innovative approaches in the field of R&D evaluation methodologies, while the value of experiments ideal are well understood in social sciences since several decades. Note though that the focus is on randomized experiments and not on natural experiments or recently proposed entrepreneurial experiments. The latter one denotes 'experiments that increase the diversity of technical, organizational and institutional arrangements in which scientific research is conducted.' *(Huang and Murray 2010:567)*. However, as will be argued later on, randomized trials may be a good instrument to foster diversity, too.

The main argument of the paper is that *(selective)* use of randomization is justified because it enables more credible estimations of causal policy effects and can improve outcomes of policy intervention due to a more 'creative' project selection compared to traditional methods such as peer review. Besides applications of social experiments in the 'real world', it is still useful to consider the mechanisms that render experiments the most credible way of evaluation because every Microeconometric approach to programme evaluation tries to emulate the experimental ideal in one or the other way. Hence, pedagogical reasons justify as well the discussion of social experiments in applied social research.

The paper proceeds as follows. The fundamental problem of any programme evaluation is presented in chapter 1, which leads directly to chapter 2 on selection bias, the problem which every serious evaluation has to take into account in one or the other way. Selection bias is a very common phenomenon in empirical economic research and the main motivation for conducting social experiments, presented in chapter 3, because they offer the possibility to remove any selection bias. Section 4 argues that randomization could improve not only the validity of evaluations but also the return on public spending on R&D. An example of a social experiment in the field of R&D evaluations is discussed in chapter 5. The final section concludes with some remarks on the political economics of experimental evaluation.

The fundamental problem of causal inference

The model used to address and analyze causal questions is the meanwhile widely applied Rubin Causal Model *(RCM)*, based on several articles from the statistician Rubin *(e.g. Rubin 1974)*. Almost all of the studies and methods in empirical micro econometrics on programme evaluation refer to the RCM. The main ingredients are the **1.** notion of potential outcomes and **2.** the differentiation along several assignment rules, which determine the way the treated are selected into a programme *(Wooldridge and Imbens 2009)*. Randomization is one of these assignments mechanisms.

According to the concept of potential outcomes, an economic agent *(household, firm)* can either participate, Di = 1, or not participate, Di = 0, in a R&D programme. Hence, D represents a binary treatment indicator. In the following, it is assumed that the respective population i = 1,…,N represents firms. Hence, there are two potential outcomes, depending on the realization of Di. Assume that Y is a continuous variable that represents some outcome measure targeted by the R&D programme, for simplicity we can think about as patent applications per time period. The realization of that variable under treatment is $Y_i(1)$, while it is $Y_i(0)$ if firm i does not receive public resources for certain R&D activities. Subtraction of the latter from the former potential outcome gives us the causal effect of the policy intervention on firm i, since the only difference between these two situations is the treatment.

$$2.1 \quad \Delta_i = Y_i(1) - Y_i(0)$$

Equation *(2.1)* truly compares the comparable and there is no superior way to detect causality. Alas, this equation is not identified, because one of the two terms on the right hand side will never be realized, since a firm can either receive public R&D subsidies or not. Consequently, it is impossible to observe the effect of D on Y for economic agent i. After Holland *(1986)*, this is labelled the 'fundamental problem of causal inference'. Note, that even a social experiment cannot generate this missing observation. The actually realized outcome of the two potential outcomes is given by

$$2.2 \quad Y_i = Y_i(1)\, D_i + Y_i(0)\,(1 - D_i)$$

This difference between potential and actual realized, observable outcomes is 'the hallmark of modern statistical and econometric analysis of treatment effects.' *(Wooldridge and Imbens 2009, 9)*. Models are developed for the potential and not solely for the realized outcomes. The impossibility to observe firms in both states of treatment, i.e. treated and non-treated, suggests that the core problem of causal questions is a problem of missing data. Another implication of *(2.1)* is that it is impossible to determine causal effects at the level of the individual economic agent. Accordingly, the necessity arises to find outcome measures Δ_i based certain averages. While there is a vast literature on different estimators of treatment effects we will consider only the ATT, which is probably the most frequently applied treatment estimator. The ATT is the appropriate evaluation parameter if participation in the programme is voluntary and there is no intention to treat the whole population, i.e. e.g. all firms in a country. The ATT measures the mean effect of the programme on the treated firms, i.e. it answers the question, if and by which scale public support causes an increase of patents under the treated firms. Note that this estimator assumes homogenous treatment effects of the programme as well as no spillovers between treated and non-treated firms or other general equilibrium effects. Microeconometric programme evaluation models are essentially built on the assumption of a partial equilibrium. This assumption might be a serious concern in innovation studies, since one of the main rationalities for public support for private R&D is the presence of positive spillovers from R&D activities *(Klette et al. 2000)*.

Yet, the problem of missing data arises again and the ATT is again not identified because we will never know the innovative performance of treated firms, had they not been treated. That is why it is called the counterfactual. However, since we have replaced individual values as in *(2.1)* with expectations, i.e. averages, we are able to find observable substitutes.

The selection problem

The fundamental evaluation has at its core a problem of missing data. Hence, the question arises as how to find a substitute for those missing data. In other words, we need a so called identification assumption to replace the counterfactual with empirically realized outcomes of some comparison group. Critically, identification assumption cannot be tested of proved; they have to be carefully argued. Without exceptions, every research strategy trying to estimate causal effects has to make some identification assumptions. That is why definite results are out of reach in policy evaluation and the social sciences in general *(Bauer et al. 2009, Schmidt 1999)*.

This substitute has to fulfil two conditions: It has to be **1.** observable to the researcher and **2.** it should be a valid substitute for the counterfactual outcome for the treated. The first condition is the easier one to satisfy. For example, the cross-section estimator uses observations on non-participants to estimate the impact of the respective policy intervention. What makes things difficult is the second condition, which demands validity of the control group. A control group is valid for a cross-section comparison if the only difference between the member of the control group and the treated group corresponds to the fact that the latter one is treated and the first one is not. Hence, difference in outcomes between them will give us the causal effect of the policy intervention. In other words: We have to compare 'apples with apples' and not something else. Unfortunately, this is nothing but easy in the case of R&D policies. Project selection processes in R&D programmes aim at selecting the 'winners', i.e. those firms with the highest probability of completing the R&D project successfully. Thereby, funding agencies create a positive selection from the total population of firms eligible for public funding. Simply comparing those who received funding with those who have not is not a valid strategy to detect causal effects of funding. The former one would presumably performed much better than the latter even

in the absence of public support. As a result, using non-supported firms as a substitute for the unobservable counterfactual situation of what would have happened to the funded firms were they not funded, produces a biased treatment estimator. The reason for this unpleasant truth is the selection process that separates funded from non-funded firms. That is why the induced bias is called selection bias.

How important is the phenomenon of the selection bias? Can we simple ignore it, because of its small size? Certainly not. Angrist and Pischke *(2009)* cite an example from health economics, demonstrating the danger of being ignorant to B. Let us assume that one has to evaluate the impact of hospital treatment on the health status of individuals. The evaluative question is simple but not as simple to answer as it might seem: Do hospitals make people healthier? Following the example in Angrist and Pischke *(2009)*, *Table 1* shows the data for individuals who stayed in hospital in the last year for at least one night, *(D = 1),* and for individuals who did not stay in hospital in the last year, *(D = 0)*. The outcome measure is the mean health status, whereby 1 indicates poor health and 5 excellent health.

This data structure corresponds to the situation of *(3.1)* and the respective identification rests on a cross-section estimator *(3.4)*.

Table 1
Do hospitals make people healthier?

Group	Sample Size	Mean Health Status	Standard Error
Hospital (Treated)	7,774	3.21	0.014
No Hospital (Non-Treated)	90,049	3.93	0.003

National Health Interview Survey 2005, cited by Angrist and Pischke (2009)

Analyzing the difference in means between the two groups in *Table 1* gives us -0.72, a highly significant difference with a t-statistic of 58.9, indicating that going to hospital makes people sick. While there might be some explanations for this unexpected outcome, we should think about the possibility of the presence of a selection bias. Simple reasoning reveals that there is a certain self-selection behind these results: Only sick people go to hospital, while healthy people do not. If sick people do not go to hospital, the might be a much larger negative difference in mean health status. Assuming that on average treatment in hospital impacts in a positive way on health status (ATT>0), which we cannot infer from the data in *Table 1*, and referring to the formulation in *(3.3),* it is clear that the positive treatment effect ATT is completely overshadowed by a large, negative selection bias, |B|>ATT. As a result, simple accounting strategies deliver fundamentally misleading answers to question on causal programme effects.

Being aware of the problem of selection bias, careful analysis will detect it nearly everywhere in society because economic agents are rational agents in the sense that they try to maximize some objective function. For example, Guadalupe et al. *(2011)* ask the research question why foreign owned firms are typically more productive than domestic-owned ones. Traditionally, this is explained with knowledge spillovers from foreign firms that improve the productivity of domestic firms after mergers and acquisitions take place. Yet, the study from Guadalupe et al. *(2011)* suggests that the superiority of foreign-owned firms is largely due to selectivity, because foreign owned firms deliberately by those firms, which are already in the most productive part of the productivity distribution of firms. This selection bias explains about two-thirds of the productivity premium associated with foreign-ownership.

Which mechanisms can invalidate the comparison of treatment and control group in the case of public R&D-programmes? To start with the comparison of participants and non-participants, purposive programme placement as well as self-selection of participants into the programme produces potentially large differences in certain variables and expected outcomes between those two groups. These selection mechanisms are further complicated in the case of R&D programmes as compared to typical labor market programmes, which serve typically as the main example in the microeconometrics of programme evaluation. Additionally to eligibility and participation decision, programme managers have introduced several, sometimes quite sophisticated selection mechanisms, to ensure that only projects with specific attributes and high chances for success are funded. This selection by the funding agency, one may call it granting selection as opposed to application and eligibility selection, is something very different from labor market programmes and complicates R&D programme evaluation substantially.

There is now solid evidence that funding agencies follow a picking-the-winner strategy in their granting decisions. By doing so, speaking technically, they purposeful produce a selection bias, thereby hampering simple comparisons of participants and non-participants in the R&D-programme. As a result, it is simply not valid to attribute differences in some innovation outcome measure between these two groups solely to policy interventions. Given the granting strategy, supported firms are also those firms who would have been the most successful even in the absence of funding. Hence, blunt comparisons and deduced treatment effects are almost certainly plagued by a positive bias, overestimating programme effects and the power of policy makers to alter innovation behaviour of the national business sector.

Taken together, the discussion of the problem of missing data leads us to the problem of sample selection. Randomization provides a methodologically sound answer to this problem.

Randomization delivers more credible answers to casual questions – under certain conditions

Social experiments derive their credibility from randomization. This simple mechanism has two advantages that render social experiments more credible than alternative identification strategies based on observational data: Firstly, they need fewer theoretical assumptions than and, secondly, they are easier to understand.

The allocation of treated and non-treated firms into the treatment and control group respectively ensures that the only remaining difference in outcomes can be attributed to programme participation. Taking the differences between them is a valid estimator for the causal effect of the programme. Robert Fisher, the famous statistician who implemented the first randomized trials to detect the impact of fertilizers on yields in the first half of the twentieth century argued convincingly that the only way to achieve equivalence between treatment and control group is random assignment of units under investigation into treatment and control group. One important reason for this is the problem of unobserved heterogeneity that plagues the calculation of causal programme effects with non-experimental data: While data typically

contain important observable variables that allow econometricians to control for differences between treatment and control group arising from these variables, they do not contain - per definition - information on unobservable variables such as motivation, intelligence, management quality etc. Hence, there remains an unobserved heterogeneity between treatment and control group which might biases our treatment effect estimator. If, for example, supported firms are also characterized by higher management quality and we do not control for this fact, out estimate of programme effects is bias upward.

Even though econometric theory provides some genius to this problem, they have to make certain, rather strong theoretical assumptions render their results fragile. Herein lies the big and most important advantage of social experiments: 'Random assignment also removes any systemic correlation between treatment status and both observed and unobserved participant characteristics. Estimated treatment effects are therefore free from the selection bias that potentially taints all estimates based on nonexperimental sources of information.' *(Burtless 1995:68).* To say it again in the most simply way: Randomization ensures that we truly compare 'apples with apples'.

Yet, despite these very strong and persuasive theoretical results, social experiments are fare from problem free and are increasingly criticized on validity grounds. According to Heckman et al. *(1999:1899)* it is now recognized, that 'social experiments, like other evaluation methods, provide estimates of the parameters of interest only under certain behavioural and statistical assumptions.' Randomized trials can be questioned on grounds of internal validity and external validity. The former criterion refers to the question if the results of the experiments can be generalized to the population to which the participants in the social experiment belong. The latter criterion asks the question if the results can be generalized to other populations. For example, if the introduction of a new R&D policy programme starts with a social experiment in one region, it should be asked, if the results of this pilot is representative for the respective pilot region and for the country as a whole. Probably this will not be the case.

Two specific mechanisms stand out that might invalidate the external and/or internal validity of social experiments. Heckman and Smith *(1995)* argue that randomized control groups have to fulfil two assumptions to represent the true counterfactual outcome for the treatment group. These two assumptions can be stated as the absence of two specific biases, which may be present in experimental settings: Neither **1.** 'randomization bias' nor **2.** 'substitution bias' is present.

The substitution bias can be a serious threat to the validity of social experiments especially in policy fields, where the number of programmes is huge and substitution between them possible. This would be the case in Austria, for example. Randomized-out firms may try to receive similar public support via other R&D support programmes. As a result, the difference in outcomes between treatment and control group cannot be attributed unambiguously to the programme under consideration. Medical trails rely on double blind designs and placebos to prevent substation bias.

Randomization bias refers to the situation in which the composition of the population participating in a programme is a function of the assignment mechanism applied. For example, it might be the case that several firms object treatment assignment via randomization and do not apply under this regime, while under traditional peer review-based assignment they would apply for funds. Hence, the results of such an experiment are not representative for the business sector as a whole.

Finally social experiments may not recover the average treatment on the treated but a so called 'intention-to-treat'-parameter. Why randomized firms do not participate in the policy programme has several reasons. In medical trails, this is a serious point of concern. The intention-to-treat analysis demands the use of every subject who was randomized according to randomized treatment assignment. Noncompliance and other deviations from the original randomized assignment are neglected. The results of an 'intention-to-treat analysis' are of course different from the causal effect of the respective drug. The argument in favour of this kind of analysis is that the results reflect the utility of a treatment for clinical practice *(LaValley 2003).* Yet, given our motivation to consider randomized trials, this may not be a satisfactory answer to the problem that we may not recover our main parameter of interest, the average treatment effect on the treated.

Why randomization might trigger even a higher return from public spending on R&D

Randomization has the potential to yield a double dividend: Firstly, it renders evaluations more credible, as demonstrated in the former sections. Secondly, it may improve project selection procedures of agencies and increase the social return of public R&D funding. 'The use of formal randomization has become more widespread in the social sciences in recent years, sometimes as a formal design for an evaluation and sometimes as an acceptable way of allocating scare resources' *(Wooldridge and Imbens 2009, 12).* The argument in favour of randomized allocation is based on the critique of standard peer-review-decision-making processes, which produces several biases in the allocation of public funds. Peer review procedures can be criticized inter alia on the following reasons:

- **Conservative bias:** Peer reviewers are prone to accept applications that confirm their theoretical or technical point of view. Because the composition of reviewers will resemble more or less a certain mainstream in the specific scientific of technological field, this will systematically prohibit the funding of innovative projects, which are by definition not part of any mainstream. They are fundamentally 'new' in the sense of Schumpeter *(2005).* Yet, as pointed out recently by Acemoglu *(2009),* technological progress displays a suboptimal diversity in equilibrium ('too much conformity') than it is optimal from a social point of view. This tendency might be fostered by a conservative bias in peer review decision making *(see also Huang and Murray 2010).*
- **Pretence-of-knowledge bias:** There is an even more fundamental argument against the use of significant public resources to identify the usefulness and commercialization possibilities of innovations. Hayek *(1974)* argues in his Nobel prize speech against the pretence of knowledge: 'if man is not to do more harm than good in his efforts to improve the social order, he will have to learn that in this, as in all other fields, where essential complexity of an organized kind prevails, he cannot acquire the full knowledge which would make mastery of the events possible.' Thinking again about the work of Schumpeter, Hayek might have a point especially in the selection of innovative projects. To give two *(provocative)* examples mind the following two quotations *(both are cited in Brezis 2007:692)*: The first stems from the CEO of IBM, 50 years ago: 'I think that the world market for computers is for no more than five computers.' The second quote stems from the commander of the allied forces in World War I: 'Planes are a nice toy but with no military value.' Delegating certain allocations of funds to the mechanisms of chance, is, considered in this way, no abdication of human rationality but shows a deeper insight into it.
- **Risk-aversion bias:** This bias is strongly interrelated with the conservative bias, but it deserves separate attention because some projects might be risky, even if they follow a well-known technological paradigm or sectorial-innovation patterns. Hence, funding of these projects might reproduce structural patterns. However, funding agencies are rather risk averse economic agents. They try to fund those projects with the highest probability of success. Following Einiö *(2009, 1),* this is no optimal strategy even from a social point of view: 'A major concern is that

programme managers may be encouraged to support projects with the best technical merits and the highest potential for commercial success. As these projects typically have high private returns they will be undertaken even in the absence of the support. In this case government support may induce only a little additional R&D if any at all.' In the same vein, Tichy *(2009)* argues for the case of Austria that it is almost certain, that most of the public funded R&D projects would have been conducted even without subventions. Linked to this argument is Tichy's diagnosis that the typical subsidized project hardly contributes to any radical innovation; the majority aims at the advancement of already well-established technologies.

Given these problems, randomization might improve on these peer-review-based outcomes. It will produce a wide variety of funded projects without any need to pretend knowledge on unknowledgeable things. Furthermore, there will be no adverse selection of risky projects and the funding agency could not be punished in the case of project failures, the almost natural outcome of funding risky innovation projects: 'In all policy making one has to accept that mistakes are being made.' *(Chaminade and Edquist 2010:111)*. Last but not least, chance is probably the cheapest way of allocating funds thereby raising efficiency levels of funding agencies.

Of course, there is no need to abandon any valuation of projects entirely. One of the proposed randomization mechanisms suggests differentiating projects into three groups *(Brezis 2007)*: **1.** projects which should be funded in any way, **2.** projects which should be funded in no way and **3.** projects which are in-between. If project quality follows a normal distribution, the latter group might well be the largest one. Randomization is only applied to this third group. This so called 'focal randomization' was already applied for two years in the allocation of funds by the ISF *(Israeli Science Foundation)*.

Evaluating the effectiveness of innovation vouchers with randomized treatment and control groups – the Dutch case

To the best of my knowledge, the evaluation of the Dutch innovation voucher is the only experimental R&D programme evaluation conducted so far. In the following, the intervention and the evaluation design is described based on the Cornet et al. *(2006)*.

The aim of the innovation voucher scheme is to increase the interaction between small and medium sized enterprises *(SMEs)* and public research organizations *(PROs)*. This meanwhile widely used policy instrument – even Switzerland intends to introduce such a scheme – addresses specific barriers for SMEs to search and use knowledge produced by public or research organizations. The Dutch innovation voucher is a credit note with which SMEs can commission an application oriented research question from a public research organization. It is worth 7,500 and cannot be cashed-in. The innovation voucher scheme was launched in the form of a pilot scheme with 100 vouchers by the Ministry of Economic Affairs in 2004. Originally, the principle of allocation should be based on 'first come, first serve'. But if the number of applicants exceeds the number of offered vouchers on a single day due to oversubscription, randomization in the form of a lottery is used instead to allocate limited resources. As 1,044 SMEs applied for 100 vouchers on the first day the pilot scheme was launched in a first round in September 2004, a lottery was applied. As a result, the total of 1044 SMEs w randomly divided into a group of 100 voucher winners and 944 voucher losers. In the language of experiments, the former serve as treatment while the latter as control group. This created the possibility for a rigorous estimation of treatment effects induced by innovation vouchers. Given the state of evaluative knowledge on the effectiveness of R&D policy in the Netherlands, summarized by Cornet et al. *(2006:9)* with the statement that 'at the moment not much is known about the effectiveness of current Dutch innovation policy', randomization seems to be a great progress towards in the strive for evidence-based policy-making.

The aim of the evaluation exercise undertaken by the CPB Netherlands Bureau for Economic Policy Analysis is to estimate the effectiveness of the programme against its objective to intensify the interaction between SMEs and PROs. Hence, the central research question is: 'What is the effect of the innovation voucher on the commissioning of assignments by small and medium-sized enterprises from public research institutions?' *(Cornet et al. 2006:11)*. Effectiveness is measured by the difference in assignments between treated SMEs, i.e. voucher winners, and non-treated SMEs, i.e. voucher losers.

Two data sources were used for the analysis. Firstly, the application form provides data on turnover, size (staff numbers), industry and region for treated and no-treated SMEs. The outcome variable is given by number of research assignments by treated and non-treated SMEs. Clearly, theoretical arguments suggest that the voucher scheme should have a positive effect on assignments. Information on this variable was collected via a telephone interview among a sample of treatment and control group. Out of the 1044 SMEs that applied for a voucher in September 2004, 600 were asked to participate in a telephone interview in May 2005. The overall response ratio was 52% and the number of respondents 313. 71 SMEs of the 313 have been allocated a voucher, 242 have not. As a result, the response ration amounted to 71% among voucher winners and 48% among voucher losers.

Table 2 shows that 71 voucher winners commissioned 62 assignments, while 242 voucher losers commissioned 20 assignments.

Table 2
Descriptive Statistics of assignment commissioning from October 2004 to 31 December 2004

Group	Sample Size	Number of assignments
Total number of firms	313	82
Voucher winners	71	62
Voucher losers	242	20

Cornet et al. 2006

As a result of randomization, neither any observable nor any unobservable variable should be different on average between treatment and control group. Indeed Cornet at al. *(2006)* indicate that observable characteristics do not differ in a significant way. The dependent variable of the estimated model is given by the probability of assignment. As was already mentioned, the lottery balances any characteristic between the two groups and there is no necessity to include any control variables, which simplifies the analysis even further.

Table 3 shows the estimates for the programme effects. The treatment variable *(β)* has a positive and highly significant treatment effect. The size of the coefficient gives us the information, that receiving a voucher increases the probability of an assignment by 79% relative to voucher losers. Adding control variables does not change the results. More interestingly, they are all insignificant because of randomization.

Table 3
Effect Estimation in the linear probability model

	Estimate	Standard error	P-value
Constant (α)	0.08	0.02	0.00
Effect of the voucher (β)	0.79	0.04	0.00
R^2	0.57		
N	313		

Cornet et al. 2006

Taken together, the results indicate a strong positive effect of the innovation voucher on the propensity of SMEs to interact with PROs. This effect can be interpreted in a causal manner, because randomization ensures that there are no other factors with explanatory power that might explain the difference between voucher winners and voucher losers.

The external validity of this social experiment has to be questioned. Again, selection problems have to be considered. There might be a self-selection of innovative SMEs into the first pilot scheme and the 100 voucher winners and 944 voucher losers under consideration might differ systematically from those SMEs which may apply for a voucher after the programme is implemented economy wide. If this is the case, than the results stated above display a positive selection bias and the causal effects for the average SME in the Netherlands should be smaller than that for the average SME in this pilot scheme. However, the potential bias in participation due to random assignment *(randomization bias)*, may be no point of serious concern, since the applicants could not know in advance which kind of assignment mechanism will be actually applied: first-come, first serve or lottery. Another potential bias comes from general equilibrium effects after economy-wide introduction of the voucher scheme. For example, university researchers might pay more attention and scrutiny to the first 100 assignments by SMEs but they become bored or even annoyed when the number of strongly applied questions from SMEs increases over a longer time horizon.

Summary and remarks on the political economy of experimental evaluation

Social experiments are a seldom applied but very powerful evaluation method. However, increasingly scare public resources and the 'credibility revolution' in econometrics triggered an increase of social experiments, especially in labour, education and development economics. While there are very few examples for randomized trials in innovation economics, there are some good arguments to rethink this situation and consider possibilities for social experiments.
This article described the 'fundamental problem of causal inference', which renders simple accounting strategies as evaluation methods completely invalid. The core problem of missing data on counterfactual outcomes follows directly to issues described by the presence of sample selection bias as the main challenge that every serious evaluation has to address. It was demonstrated that randomization provides theoretically a perfect solution to this problem. Yet, substitution bias and randomization bias may hamper the internal and external validity of experiments.

Besides the power of experiments to solve the sample selection problem, randomization can even improve the social rate of return on public R&D-spending due to an improvement of traditional allocation mechanisms. Peer review is plagued by three biases: A conservative bias, a pretence-of-knowledge bias and a risk-aversion bias. Some of these biases can be improved due to the selective application of randomization. It was also stressed that randomization is quite cheap compared to professional peer review processes.

Finally it is important to say something about the political economy of social experiments: From a pure theoretical point of view it is questionable, why there are so few randomized trials, leaving the effectiveness of several state interventions a widely unknown question. There are of course ethical concerns, but one wonders why sick people in developed countries and absolute poor people in developing countries are more 'qualified' for randomized trials than e.g. firms. A political economic explanation is necessary to explain this contradictory situation. Firstly, policy makers are reluctant to conduct social experiments because the results are so credible that a reinterpretation in their own favour is severely limited, at least compared to non-experimental results *(Schmidt 2007)*. Secondly, a similar argument can be applied for programme managers because current allocation procedures give them much more leeway *(Giebe et al. 2006)*. Yet, giving up this power is obviously no way of utility maximization, as theories on the behaviour of bureaucrats suggest.

Literature

Acemoglu D(2009)
Diversity and technological progress. NBER Working Paper, No. 16984.
Angrist J, Pischke JS (2010)
The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. In: Journal of Economic Perspectives, 2, 3-30.
Angrist J, Pischke JS (2009)
Mostly harmless econometrics. An empiricist's companion. Princeton.

Bauer T, Fertig M, Schmidt C (2009)
Empirische Wirtschaftsforschung. Berlin, Heidelberg.
Brezis E (2007)
Focal randomization: an optimal mechanism for the evaluation of R&D projects.
In: Science and Public Policy, 10, 691-698.
Burtless G (1995)
The case for randomized field trials in economic and policy research.
In: Journal of Economic Perspectives, 2, 63-84.
Cerulli G (2010)
Modelling and Measuring the effect of public subsidies on business R&D:
A critical review of the econometric literature. In: Economic Record, 421-449.
Chaminade C, Edquist C (2010)
Rationales for public policy intervention in the innovation process: systems of innovation approach.
In: Smits R et al. (eds.): The theory and practice of innovation policy. Cheltenham, 95-114.
Cornet M et al. (2006)
Do innovation vouchers help SMEs to cross the bridge towards science? CPB Netherlands Bureau of Economic Policy Analysis Discussion Paper, No. 58.
Cornet M, Webbink HD (2004)
Towards evidence based policy. ? CPB Netherlands Bureau of Economic Policy Analysis Discussion Paper, No. 48b.
Czartnitzki et al. (2003)
Einsatzmöglichkeiten neuer quantitativer Ansätze zur Wirkungsanalyse öffentlicher Forschungsförderung –
Eine Grundlagenstudie. Mannheim.
Einiö E (2009)
The effect of government subsidies on private R&D: evidence from geographic variation in support funding.
Helsinki Center of Economic Research Discussion Papers No 263.
Giebe T et al. (2006)
How to allocate R&D (and other) subsidies: An experimentally tested policy recommendation.
In: Research Policy, 1261-1272.
Hayek, FA (1974)
The pretence of knowledge. Nobel prize lecture.
www.nobelprize.org/nobel_prizes/economics/laureates/1974/hayek-lecture.html, 25.11.2011.
Heckman J et al. (1999)
The economics and econometrics of active labour market programs.
In: Ashenfelter A, Card D (eds): Handbook of labor economics, Vol 3, 1865-2097
Heckman J, Smith J (1995)
Assessing the case for social experiments.
In: Journal of Economic Perspectives, 2, 85-110
Holland P (1986)
Statistics and causal inference. In: Journal of the American Statistical Association, 945-960
Huang K, Murray F (2010)
Entrepreneurial experiments in science policy: Analyzing the Human Genome Project.
In: Research Policy, 567-582

Imbens G, Wooldridge J (2009)
Recent developments in the econometrics of program evaluation.
In: Journal of Economic Literature, 1, 5-86.
Jaffe A (2002)
Building programme evaluation into the design of public research-support programmes.
In: Oxford Review of Economic Policy, 1, 22-34.
Klettet et al. (2000)
Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies.
In: Research Policy, 471-495.
LaValley, M (2003)
Intent-to-treat Analysis of Randomized Clinical Trials
people.bu.edu/mlava/ITT%20Workshop.pdf, 25.11.2011
Link A, Scott J (2011)
Public goods, public gains. Calculating the social benefits of public R&D. Oxford.
Polt W, Pointner W (2005)
Diffussionsorientierte Technologiepolitik. Eine vergleichende Wirkungsanalyse für Österreich, die Schweiz, Deutschland und die USA. Graz.
Republik Österreich (2011)
Der Weg zum Innovation Leader. Strategie der Bundesregierung für Forschung, Technologie und Innovation.
RTD-Evaluation Toolbox (2002).
Rubin D (1974): Estimating causal effects of treatments in randomized and nonrandomized studies.
In: Journal of Educational Psychology, 5, 688-701
Schumpeter JA (2005)
Kapitalismus, Sozialismus und Demokratie. Tübingen.
Schmidt C (2007)
Policy Evaluation and Economic Policy Advice. Ruhr Economic Papers, No. 1.
Schmidt C (1999)
Knowing what works. The case for rigorous program evaluation. IZA Discussion Paper Series, No. 77.
Streicher G (2007)
Additionality of FFG funding. InTeReg Working Paper No 49.
Tichy G (2009)
Was ist das 'Österreichische' an der österreichischen FTI-Politik?
In: Leitner et al. (eds.): Innovationsforschung und Technologiepolitik in Österreich. Neue Perspektiven und Gestaltungsmöglichkeiten. Innsbruck et al., 255-272.

Author

Christian Reiner
Joanneum Research
MAIL christian.reiner@joanneum.at

Peter Kaufmann and Sabine Mayer

Is there a use for counterfactual impact evaluation?

1

Introduction

Capturing impacts of policy interventions is receiving increasing attention in the EU. Because of budgetary constraints, the allocation of public means needs to be argued on the basis of sound evidence.

Monitoring systems have traditionally been laid out to measure what could be measured easily, i.e. inputs, outputs and results. But impacts are what policy makers and the public in general are ultimately interested in. The resulting data constraints made it often only possible to apply relatively simple evaluation methodologies that are not capable of measuring impacts in a plausible manner. Others tried to capture impacts with a more encompassing approach, but had to base their models on questionable assumptions due to lacking data to produce reliable estimates.

A plea for using different methodological approaches for impact evaluation

In the history of policy evaluation, a range of methodologies have emerged due to different epistemological beliefs prevalent in various research communities. The discussions have ranged around more or less theory-based frameworks, what to include in these theories, how to capture positive and negative side effects on micro and macro levels, whether to apply qualitative or quantitative methods, external versus internal evaluations, etc.

In the meantime, it is generally acknowledged that it is important to put an evaluation on theoretical footings. This is especially true for impact evaluations where lacking theory would increase the risk that arbitrary and even wrong results cannot be separated from true effects because no hypotheses testing would occur. Here, a distinction can be made between traditional 'theory-driven evaluation' and 'realist(ic) evaluation' *(Pawson and Tilley 1997)*. The traditional approach *(building on Karl Popper's positivist approach on deriving hypotheses and putting them to test in a 'piecemeal social engineering' fashion)* favours *(quasi-)*experimental settings to minimize biases accrued from the fundamental evaluation problem of non-observability of the counterfactual. The realist approach challenges this view because the mostly econometric methods applied in this tradition miss the focus on the mechanisms that make an intervention work as well as under-represent varying contexts that can influence the success and the inner logics of public interventions decisively. This approach reframes the evaluation questions *(does it work? …)* into 'what works for whom under which circumstances?' Thus, the main purpose of evaluation becomes to 'find out how and under what conditions a given measure will produce its impacts' instead of 'shall the public funding be continued *(in this way)* or redistributed to different purposes'.

[1] This paper draws partly on Kaufmann, P. Henning, H. C. A., and J. Michalek, 2009 'Comparing methods for the evaluation of EU rural development programmes', ADVANCED EVAL working paper series, Kiel.

Given the two basis impact evaluation questions, **1.** To what extent does the policy work? and **2.** Why and how does the policy work?, quantitative evaluations following more 'traditional' theoretical approaches will give indications on the first evaluation question, but need to be complemented by other, mostly qualitative approaches to substantiate the above and give answers to question two. Advocates of the realist evaluation approach would probably argue that a further question needs to be added **3.** How can the programme be improved?, and that this approach tries to give answers to all three questions, mostly in a qualitative way, more tailored to the specific circumstances.

Whatever research tradition one decides to follow, one can state that it is getting increasingly common for impact evaluation to draw on some form of programme theory *(logic model, intervention logic) (Leeuw and Vaessen 2009; Lipsey and Cordray 2000; World Bank 2006)*. Still, the methodology used for uncovering outcomes vary considerably between research traditions, although there is an increasing call by international organisations for 'rigorous impact evaluations', meaning the application of quantitative methods to estimate net-impact and minimise selection bias.

At least during the last 20 years, public support has developed into rather broad and complex programmes, embedded in policy measures that can influence each other, where individual measures show different causal chains and overlaps *(and potential trade-offs)*. Thus, evaluators have increasingly been forced to apply a mix of methods to answer multiple evaluation questions, though they were struggling with how and what to combine and integrate to get coherent results. With time, researchers tried to turn this 'vice' into a 'virtue' by establishing a mixed-method research programme. Here, the aim is to combine the strengths of individual theories, methods, and/or data types to reach a more encompassing and useful picture. In the optimal case this means that different methods or types of data are combined to answer a common research question, which is interdependent, either implemented sequentially or simultaneously. This poses the challenge of fruitful collaboration across disciplines or acquiring new skills. But there are advantages to the combination of approaches due to the complexity of today's breath of policy interventions, which is increasingly recognized in the field *(e.g. Leeuw and Vaessen 2009)*.

The integration of qualitative and quantitative methods to answer a research or evaluation question can be conceptualized as follows: The intervention logic sets the frame for the evaluation of singular measures on the micro level by conceptualizing the causal chain from inputs to impacts. Qualitative methods are necessary to put the theoretical construct on firm footings through covering background knowledge to formulate correct quantitative models and also potential alternative conceptualisations/modifications to the intervention logic. Quantitative methods are advantageous for estimating and comparing the magnitude of net-impacts. Qualitative methods can again be used to gain a deeper understanding of why the quantitative effects played out the way they did and which role heterogeneity in implementation and context plays in determining the success of policy measures.

The full impact cannot be identified by definition because it cannot be observed directly. This is why it needs to be approximated by some technique to minimise selection bias[2]. How this is done necessitates a good understanding of the domain and how beneficiaries are selected, which determines how biases may be generated. Minimisation of selection bias is normally done by establishing a counterfactual sample with which beneficiaries can be compared to.

[2] Selection bias can originate either because policy programmes select a particular target group by assigning eligibility criteria to a subsidy, or because a policy programme attracts a particular sub-set of the population due to some unforeseen reason *(e.g. larger businesses tend to be better informed and their propensity to apply for support might be higher; or take up might differ simply because the quality of advice of support services differs)*.

For being able to judge what individual methods can contribute to impact evaluation, we discuss in the following some basic evaluation problems which are then used as criteria to assess the individual methods presented one by one.

Why is counterfactual analysis important?

Although it is not possible to observe what would have happened in the absence of the policy, one can still approximate this situation by constructing a baseline through comparing beneficiaries with non-beneficiaries. To do this in a credible way one also needs to minimise selection bias. Thus, the two groups should have very similar characteristics that can be tested statistically. 'The appropriate method for dealing with selection bias depends on the nature of selection into the treatment group, the particular evaluation problem and the richness of the data. Different methods for dealing with selection bias and the associated estimators will only yield unbiased estimates if the assumptions underlying these methods are in fact correct. For these reasons it is important to carry out as thorough an assessment of these assumptions as is possible' *(Riley, et al. 2007: 45)*.

A further challenge in uncovering the real impact of a policy measure is given because there are nearly always exogenous factors that also influence the development outcomes of the units of analysis. If, for example, companies receive funding to carry out specific projects or for their start-up phase, an impact assessment should help to disentangle effects of a given programme from effects of other exogenously determined factors.

Evaluations in general, thus also quantitative approaches produce biased assessments if they miss to include all relevant influences/variables. In econometrics, this is called the problem of unobservables, for which methods have been developed to correct for (at least) partly.

The final basic evaluation problem is that every policy measure produces unintended side effects. Negative side effects on the micro level could mean that the policy did not have the envisaged impact on behaviour change: i.e. beneficiaries receive the support for something they would have done in any case *(deadweight effect)*. On a macro/meso level, resources used for one group might have detrimental effects on the non-eligible group *(substitution effect)*. Or investments in one region could displace investments in another region (displacement effect). Positive side effects could manifest themselves directly as leverage for additional private investment *(micro)* or indirectly as income multipliers (macro).

Non-experimental approaches

Ex-post analysis of beneficiaries

A simple ex-post analysis restricted to beneficiaries focuses on a retrospective reconstruction of policy influences. This is often implemented by asking beneficiaries during interviews what the influences of particular policy interventions were. Other qualitative approaches applied are expert interviews, focus groups, the analysis of monitoring data of beneficiaries on the micro or on programme level, or descriptive analyses of macro data from statistical offices or similar. The advantage of this approach is on the side of data availability because evaluators will always be able to receive the data for this kind of analysis either from own surveys or from the monitoring system. The disadvantage of this approach is that one cannot really judge the net-impact of the policy because effects of other factors are not dealt with and/or are assessed by experts/evaluators themselves qualitatively.

The following table gives a succinct overview to what extent this method helps to give answers to the basic problems evaluators' face, which were already discussed above.

Table 1
Capacity of the ex-post analysis of beneficiaries to answer basic impact evaluation problems.

Basic impact evaluation problems	Extent of incorporation in method
Minimisation of selection bias	No
Offers a solution for the omission of influences/ unobserved variables	Not in a quantitative sense. In a qualitative sense, evaluators sometimes try to minimise this bias by devising a broad evaluation framework with interactive elements to lessen the danger
Disentangles support from exogenous factors at micro level	Qualitative expert assessment, if at all
Valuation of non-traded goods	Sometimes qualitative assessment, but sometimes not incorporated or incomplete
Calculation of programme side effects (deadweight, substitution, displacement, income multiplier)	Micro effects sometimes inferred from questionnaire items or expert judgement, macro effects often not assessed

Simple before - after approach

A second relatively simple method is to compare how an indicator changed during an intervention where the beneficiaries are compared with themselves before and after the intervention had happened. This necessitates baseline data for beneficiaries to act as controls, though also here no comparison with a counterfactual group is envisaged. The advantage of before - after comparisons is that no selection effect occurs because the same unit of analysis is compared with itself. The disadvantage is that external factors influencing the outcome of interest *(e.g. macroeconomic trends or other subsidy schemes implemented either at the same time or before)* cannot be separated from the impact of the intervention. A simple before – after comparison is sufficient if there are no credible influences besides the intervention, which sometimes happen in the case of basic physical infrastructure investments. The before-after research design is sometimes somewhat improved if a longitudinal data series is available before, during and after the intervention, where evaluators try to establish whether substantial shifts had happened at some point which can be attributed to the policy intervention with some probability. But this is often still not a rigorous design because the effect of other factors is not systematically reduced.

Table 2
Capacity of the simple before - after approach to answer basic impact evaluation problems.

Basic impact evaluation problems	Extent of incorporation in method
Minimisation of selection bias	No
Offers a solution for the omission of influences / unobserved variables	Yes, because units of analysis are compared with themselves before the intervention
Disentangling support from exogenous factors at micro level	Qualitative expert assessment, if at all
Valuation of non-traded goods	Sometimes qualitative assessment, but sometimes not incorporated or incomplete
Calculation of programme side effects (deadweight, substitution, displacement, income multiplier)	Micro effects sometimes inferred from questionnaire items or expert judgement, macro effects often not assessed

Quasi-experimental designs

Because the real impact cannot be observed by its very nature, it is generally acknowledged in the evaluation literature that impact evaluations need to construct a counterfactual case *(equivalent control groups)* to get closer to the true impact of interventions. The assumption is here that the impact of a public intervention can be measured if the evaluator can compare 'on average' similar groups of beneficiaries and non-beneficiaries, and thus lessens the selection bias problem. The main challenge for quasi-experimental methods – besides data availability - is to identify all observed and unobserved characteristics of beneficiaries *(and their comparison group)* relevant for potential changes caused by a policy intervention. In general, the strategy should not be to give in to current *(i.e. less than optimal)* data situations, but instead to push for improving monitoring systems so that they are able to make before-after and with-without comparisons possible.

Selection on observables

Non-equivalent group designs
Non-equivalent group designs entail the comparison of average performance indicators of beneficiaries with the average of non-beneficiaries, i.e. the individual units are not matched. The advantage of this design is that the sample sizes needed for non-beneficiaries can be relatively small. The disadvantage of non-equivalent group comparisons is that selection bias is present because one compares beneficiaries with potentially non-comparable non-beneficiaries *(they are neither randomly sampled in advance like in randomized controlled trials, nor matched ex-ante).* As usual, this can

be implemented either as a cross-section design only using ex-post data or as a before-after comparison, where the disadvantage of cross-section estimates is that a substantial selection bias can affect obtained results.

The single difference comparison at one point in time is sometimes found when beneficiaries and non-beneficiaries are surveyed and compared. A matching approach is not implemented because the costs of enlarging the non-beneficiary sample are deemed to be prohibitive. It is sometimes also constructed from existing databases as long as there are identifiers for beneficiaries for the support programme in question. This can happen when using databases that are initially built for other than evaluation purposes in mind, but from which samples of beneficiaries and non-beneficiaries can be constructed.

Table 3
Capacity of the non-equivalent group design to answer basic impact evaluation problems.

Basic impact evaluation problems	Extent of incorporation in method
Minimisation of selection bias	No
Offers a solution for the omission of influences/ unobserved variables	Only if implemented in a before-after design
Disentangles support from exogenous factors at micro level	Qualitative expert assessment, if at all
Valuation of non-traded goods	Sometimes qualitative assessment, but sometimes not incorporated or incomplete
Calculation of programme side effects (deadweight, substitution, displacement, income multiplier)	Sometimes qualitative estimates; or rough estimates of e.g. deadweight effects are made

Group matching
Comparison group matching methods try to reconstruct ex post an experiment by choosing a group from the eligible population that should ideally only differ from the beneficiaries in one major way – that they are non-beneficiaries. One major assumption is that all relevant characteristics can be observed/measured. Here we can distinguish between exact matching and propensity score matching. Exact matching of beneficiaries and non-beneficiaries uses only conceptually important variables to correct for selection bias. The matching exercise establishes whether members of the two groups have very similar characteristics on a one-to-one basis. Here we got a dimensionality and a time problem: While the number of observable characteristics in the group of programme participants increases linearly, the number of necessary observations in the control group increases nearly exponentially because one needs a number of cases to identify exact matches. This is also a time consuming process. Propensity score matching *(PSM)* overcomes these problems by matching beneficiaries and non-beneficiaries on their conditional probability to become a programme beneficiary given observed characteristics. Rosenbaum and Rubin *(1983)* showed that when outcomes are independent

of programme participation conditional on characteristics Z, they are also independent of participation conditional on the propensity score, Pr (D=1|Z). In contrast to exact matching, in PSM, beneficiaries and non-beneficiaries are matched on the basis of only one value (i.e. the propensity score). The estimation of the propensity scores involves a logit or probit regression model with the observed decision of programme participation as dependent variable and all available covariates as independent variables *(Heckman, et al. 1997).* The resulting propensity scores are used for matching beneficiaries with non-beneficiaries. The difference in performance between the beneficiaries and control groups gives an average estimate of the programme impact.

Compared with ordinary least square *(OLS)* regressions, PSM does not impose any linear relationships between variables or a homogeneous additive treatment effect. Thus, if PSM is implemented well, it will likely result in less biased estimates in comparison with OLS. This is also the case because researchers only rarely incorporate higher-order and interaction terms in OLS regression, which would make the estimates of the two methods more similar *(Smith 2000).* Thus, PSM is more robust, but it is also rather data hungry which needs a relationship of beneficiaries to non-beneficiaries of around 1:4 to 1:10 depending on the heterogeneity of the two groups *(Kaufmann and Pufahl 2009).* The larger the sample the more likely is the probability that matched samples will be found.

While matching can also be done on cross-sectional data, matching performed prior to a given policy intervention *(e.g. PSM)* combined with a longitudinal approach *(difference-in-difference - DID)* significantly increases the quality of the analysis. The PSM-DID method compares beneficiaries and non-beneficiaries before and after an intervention, and this way controls for *(only time-invariant)* unobservable variables that also determine participation *(described below in more detail).*

A downside of PSM is that a fairly large pool of especially non participants is necessary from which adequate matches can be selected. Thus, a broad take-up of an intervention can make the measurement of the counterfactual impossible because too few credible units could be left as non-beneficiaries to form the control group. It also needs a sizeable amount of high quality variables to ensure a meaningful selection on observables. Although biases are very likely reduced in comparison to other regression methods, some kind of bias will still remain because of non-captured context variables or measurement errors in variables. For example, Diaz and Handa *(2006)* report that already small differences in the way outcomes are measured can lead to bias. Dehejia *(2005: 355)* addresses this problem by formulating that 'a researcher should always examine the sensitivity of the estimated treatment effect to small changes in the propensity score specification; this is a useful diagnostic on the quality of the comparison group'. Peikes, et al. *(2008)* caution social researchers by reporting a PSM application under 'seemingly ideal circumstances' and passing of all statistical tests, which still produced a biased impact assessment compared to the result of a randomized design. Thus, one should not overestimate the certainty of findings but instead make statistical tests and combine this promising approach with other quantitative *(especially DID)* and qualitative approaches *(to confirm quantitative findings and to get a deeper understanding of processes including feedback loops).*

Table 4
Capacity of group matching to answer basic impact evaluation problems.

Basic impact evaluation problems	Extent of incorporation in method
Minimisation of selection bias	Yes
Offers a solution for the omission of influences/ unobserved variables	Not per se – only if combined with DID
Disentangling support from exogenous factors at micro level	Yes, if respective control variables for other programme funding and funds from previous programming periods are included
Valuation of non-traded goods	If cases are not only matched along their socio-economic, but also by taking into consideration their natural and spatial characteristics
Calculation of programme side effects (deadweight, substitution, displacement, income multiplier)	Possible for deadweight, substitution, displacement, and leverage effect. No for income multiplier

Because of the practical problems with data availability, the assessment of net-effects often falls short and stays at the stage of measuring gross-effects, which are then sometimes complemented by qualitative assessments of indirect negative *(deadweight, displacement, or substitution)* and positive effects *(seldom income multipliers, hardly ever leverage).* It is argued that the data availability situation often makes it impossible to generate the datasets necessary for such an analysis within the budget and time available for an evaluation exercise. This lack of information mostly hinges on the availability of data for non-supported groups because there is no obligation to submit the necessary information on non-supported entities into a programme monitoring system.

Selection on unobservables

Difference-in-differences

In a difference-in-differences *(DID)* approach, a group of beneficiaries is compared with a group of eligible non-beneficiaries before and after a programme is implemented. This is in contrast to a simple before-after comparison where beneficiaries are only compared with each other. DID, combined with a regression approach like PSM, is generally seen as the state-of-the-art method to reduce bias because it combines the advantages of before-after and with-without designs. Basically, it can be combined with any approach that uses comparison groups before the policy implementation, which are then included in the DID estimation. Of these approaches, DID is mostly combined with the relatively robust PSM *(logit)*. This is done because both implemented on their own have weaknesses that can be reduced by combination: PSM only selects on observables, potentially leaving out unobserved characteristics and their influence over time, and standard DID only selects on unobservables, leaving out a correction for selection bias.

Table 6

Capacity of difference-in-difference approach to answer basic impact evaluation problems.

Basic impact evaluation problems	Extent of incorporation in method
Minimisation of selection bias	Yes, if combined with PSM
Offers a solution for the omission of influences/ unobserved variables	Yes
Disentangling support from exogenous factors at micro level	This happens if combined with PSM
Valuation of non-traded goods	Possible, if combined with a regression technique and applied as discussed under PSM.
Calculation of programme side effects (deadweight, substitution, displacement, income multiplier)	Possible for deadweight, substitution, displacement, Possible, if combined with PSM approach (except income multiplier)

Instrumental variables

The instrumental variable approach estimates causal relationships in a parametric regression framework. In contrast to matching, where the matching variables should affect both the policy intervention and the outcome, the instrumental variable approach relies on an exclusion restriction that requires 'a variable that determines participation in the programme but not the outcome of the programme itself' (*Blundell and Costa Dias 2000: 429).* Otherwise we would have an endogeneity problem with biased estimates. Endogeneity means that a third variable causes two other variables to correlate without them being causally connected in reality. To put it differently, an instrumental variable is not part of the original equation to predict programme participation; by including it in the regression it helps to lessen the bias resulting from the spurious correlation between two variables originally included in the estimation. Like DID, instrumental variables can be integrated in different regression approaches.

An advantage of instrumental variables is that they control for the selection of unobservables *(unmeasured factors)*, which is particularly useful when working with databases set up for different purposes. A frequent problem is that good instruments are hard to find *(given the data available from secondary data sources)* and difficult to validate. Thus, a good instrument should be agreed by a group of subject matter experts and not be an individual researcher.

Table 7

Capacity of instrumental variables approach to answer basic impact evaluation problems.

Basic impact evaluation problems	Extent of incorporation in method
Minimisation of selection bias	To what extent IV and PSM can be combined is still under discussion. See Bhattacharya and Vogt (2007).
Offers a solution for the omission of influences/ unobserved variables	Yes
Disentangling support from exogenous factors at micro level	Possible
Valuation of non-traded goods	If considered in the regression approach
Calculation of programme side effects (deadweight, substitution, displacement, income multiplier)	Possible (except income multiplier)

Selective control designs described above are advantageous in a quasi-experimental design. Still, controlling for threats for internal validity *(which is all about making sure that causal relationships do exist)* is of utmost importance and requires experience by the evaluator of the subject area, the support programme and data availability. Social experiments *(also called randomised control trials or random assignment studies)* avoid these problems altogether which is why they are sometimes called the gold standard in impact evaluation. This involves the random selection of beneficiaries and non-beneficiaries from the eligible population into two groups in advance of the policy implementation. Because these two groups will have by definition similar characteristics, it would suffice to simply compare average effects ex-post to uncover the true impact. It is argued that the most effective implementation of experiments are during the introduction of new programmes, when stakes are high, and when there is a controversy about programme effectiveness, or when policy change is desired *(Posavac and Carey 2007).* Because of the associated high costs and the ethical argument that policy makers cannot withhold a policy to some parts of the population simply because one wants to do an experiment, it is mostly not considered for impact evaluation notwithstanding its theoretical advantages. But in some instances, it is not possible to roll out a policy to all potential beneficiaries in a short time frame which leaves room for a randomized

scaling up of an intervention that creates control groups in a 'natural' way *(called pipeline approach)*. Another method, that comes close to a randomized design, with fewer obstacles, is the following discontinuity analysis.

Discontinuity analysis

The regression discontinuity analysis takes advantage of programmes that have a cut-off point regarding who receives the subsidy. Thus, one can implement this method if the eligibility for a programme is determined by some more or less artificial *(continuous)* threshold *(e.g. age, income, time, number of people, ranking of firms)*. The advantage of this method is that there is a situation that resembles randomization in the neighbourhood of the threshold. The units around the threshold experience sharply different treatments although they have very similar values for the selection variable. This is why some researchers suggest this method for getting as close as possible to pure randomization. An important restriction to its applicability is that it needs a 'sharp' discontinuity in treatment around the threshold, which is often not the case. Further, a small number of observations around the threshold results in low statistical power. This leads researchers often to extend the band around the threshold to include more units. But these units are getting more dissimilar, thus introducing bias. So we are clearly getting in some trade-off situation. A drawback can also be that the impact of the intervention in only identified 'locally'. If we envisage a situation where the programme impacts vary considerably with e.g. the firm size, it will be difficult to impossible to extend the estimated impacts around the threshold to the entire population of firms[3].

Table 7

Capacity of discontinuity approach to answer basic impact evaluation problems.

Basic impact evaluation problems	Extent of incorporation in method
Minimisation of selection bias	Yes, by comparing similar units around a threshold
Offers a solution for the omission of influences/ unobserved variables	Yes
Disentangling support from exogenous factors at micro level	Possible
Valuation of non-traded goods	Possible
Calculation of programme side effects (deadweight, substitution, displacement, income multiplier)	Possible (but not macro effects)

[3] Interested readers shall be referred to the very accessibly written introduction to the method at www.socialresearchmethods.net/kb/quasird.php

A favourite method for impact evaluation does not exist, because it depends on the circumstances and data availability for certain evaluations. Still, there are methods that should be preferred over others if the situation allows for it. The above, short overview of some of the methods applied at the micro level, should give some indication of their applicability. A different set of methods would apply in case of major funding schemes, where also macro aspects should be captured.

We now turn to our case study which applied a combination of qualitative methods, together with propensity score matching combined with difference-in-difference, to calculate the net-programme effects at micro level *(Mayer et al. 2011)*.

Case study

The Austria Business Service *(aws Austria Wirtschaftsservice)* is responsible for managing several funding schemes on behalf of the Austrian Federal Ministry of Economy, Family and Youth. In 2011, the specific funding instrument of guarantees based on the so called 'KMU-Förderungsgesetz' was evaluated by the Austrian Institute for SME Research. These guarantees are applied in a set of six funding programmes, all of them addressing SMEs. Some focus on specific phases in the development of an enterprise *(start-up phase, crisis,…)*, on entrepreneurial functions such as innovation, on the availability of financial means for small scale projects, or to raise equity. One of the programmes *(SME guarantees)* addresses access to finance for SMEs in general. Two of the programmes can combine guarantees with grants. The evaluation aimed at investigating the design, the management and the impacts of these guarantees as a 'set of funding instruments fostering access to finance for SMEs' rather than investigating the single programmes individually.

Since the tasks set out for the evaluation were broad, the methodological approach was designed to mirror the wide range of research questions. It consisted of qualitative as well as quantitative methods. The qualitative methods included the analysis of the programme documents, semi-standardized qualitative interviews with stakeholders, experts and with a defined group of experts from the funding-service units of commercial banking institutions. These methods addressed mainly research questions associated with the design and the process, and the logic functioning of the guarantees. An important aspect was the institutional setting employed for the implementation of guarantees as interplay between entrepreneur, banking institution and aws. The set of descriptive and quantitative methods applied consisted of an in-depth analysis of aws´ monitoring data, a company survey of the beneficiaries of at least one of the guarantees during the last four years and a counterfactual analysis. We focus now on the latter.

The basis for the counterfactual analysis was twofold: the Austrian Institute for SME Research receives regularly anonymous balance sheet data of Austrian companies from banks, including the aws. Based on this, we were able to identify the balance sheet data of the group of beneficiaries. Then we matched these data with other Austrian firms from the balance sheet database using the propensity score.

For the implementation, we analysed firstly the differences between beneficiaries and non-beneficiaries. This group is still likely to include beneficiaries of other funding schemes. However, this holds also true for the group of beneficiaries.

The second step was the identification of net impacts by a group matching approach. The matching was based on a propensity score calculating the propensity for each company to participate in the programmes.

The first finding in this respect is the identification of the effect of different structural variables on the propensity score. This analysis is based on the structural variables a company displays in the year before the funding proposal is accepted. The following criteria showed statistically significant effects on participation: the companies´ age, companies´ location in urban vs. rural areas, and sector. ÖNACE code D *(production of goods)* showed a positive effect on participation while code E *(Supply of Energy/Water)* and I *(Transportation, Telecommunication)* had a negative effect.

Based on these findings the propensity score for each single company in the two groups was calculated. Pairs with the least differences in propensity were matched within a group of ÖNACE codes. Obviously, only pairs where balance sheet data were available for the same period of time could be taken into account. The twin group of beneficiaries and non-beneficiaries consisted of 142 matches. Very young companies, where balance sheet data are rarely available, are underrepresented compared to their share in the guarantee scheme.

The net effects were calculated by combining the propensity score matching with a difference in differences approach. The analysis started in the year before the guarantees were granted and included three years after the projects were initiated. The results show clearly that beneficiaries had higher investment rates than non-beneficiaries, and beneficiaries experienced higher growth of turnover over the following three years. More effects should be visible in a longer time horizon.

A more detailed analysis within the group of participants revealed tentative indications of effects on employment in cases where the funded investment comprised a relatively high share of the company's total capital. These findings supported the estimation of intended effects according to the 'programme logic' established by the above mentioned set of qualitative methods. The institutional setting for programme implementation limits the need for the companies to be informed about the different programmes, it leaves a considerable share of programme-communication to the commercial banks and thus decreases programme governance on the side of aws. The decision which programme is adequate for the specific situation of the company is more or less driven by the funding experts of the commercial bank and evaluated by the experts in aws. This is especially true for those programmes that do not have specifically defined target groups. Where target groups are defined more narrowly (e.g. founders or young enterprises) aws implements targeted communication and is supported by intermediary organisations *(such as the chambers of commerce)*.

The general logic of intervention for the guarantees – to provide access to finance in cases where the private market is not willing to take the risk – corresponds with the fact that the financial situation of participants before the guarantee is in general weaker than for the average non-participants.

From the perspectives of the beneficiaries, the main effect of the guarantees on the planned project/investment is to allow for its implementation, while the main effects of the investment itself were often seen in securing turnover and employment. This is in line with the net effects that could be established.

Conclusions

There is no single best method for impact evaluation, which is very much influenced by data availability, the scale where the programme is implemented, but also the range of evaluation questions that need to be answered *(Blundell and Costa Dias 2000)*. Because of the general complexity of the programmes, the specific situation on data availability in different institutional settings, good evaluations always apply a mix of methods that, taken together, try to answer the range of evaluation questions posed at the outset. For being able to deliver such an evaluation, evaluators need to have a comprehensive overview and partly also in-depth knowledge of qualitative and quantitative methods. For more technical applications, they will sometimes have to rely on specific method experts and include them in their evaluation team.

Thus, the combination of qualitative and quantitative methods in a triangulate fashion is recommended to ensure a high quality of impact evaluations in the future. To understand the impact, one has to establish a sound understanding of the programme logic and of the mechanisms at work during implementation of the programmes, which can only be gained by qualitative methods comprising all relevant perspectives/contexts of the programmes that are evaluated. Quantitative methods, as discussed here, are clearly advantageous to capture the magnitude and direction of effects. It is then the role of as the adequate combination of methods to test and correct for potential biases and present and interpret the results in a suitable context.

References

Bhattacharya, J. and Vogt, W. B. 2007
'Do Instrumental Variables Belong in Propensity Scores?' NBER Technical Working Paper No. 343.
Revised September 2009.
Blundell, R. and Costa Dias, M. 2000
'Evaluation Methods for Non-Experimental Data', Fiscal Studies 21(4): 427-468.
Bundesgesetz über besondere Förderungen von kleinen und mittleren Unternehmen (KMU-Förderungsgesetz), BGBl. 432/1996, idF BGBl. I Nr. 130/2002
Dehejia, R. 2005
'Practical propensity score matching: a reply to Smith and Todd', Journal of Econometrics 125(1-2): 355-364.
Diaz, J. J. and Handa, S. 2006
'An assessment of propensity score matching as a nonexperimental impact estimator –
Evidence from Mexico's PROGRESA program', Journal of Human Resources 41(2): 319-345.
Heckman, J. J. and Smith, J. A. 1995
'Assessing the Case for Social Experiments', Journal of Economic Perspectives 9(2): 85-110.
Heckman, J., Ichimura, H. and Todd, P. E. 1997
'Matching as an econometric evaluation estimator: evidence from evaluating a job training program',
Review of Economic Studies 64: 605-654. Jungunternehmer- und Innovationsförderung für KMU – Haftungsübernahmen; Richtlinien des Bundesministeriums für Wirtschaft, Familie und Jugend im Einvernehmen mit dem Bundesministerium für Finanzen, in der Fassung vom 19. Mai 2009.

Kaufmann, P. Henning, H. C. A., and J. Michalek, 2009
'Comparing methods for the evaluation of EU rural development programmes',
Advanced Eval working paper series, Kiel.
Kaufmann, P. and Pufahl, A. 2009
'Contribution to the EC guidelines for evaluators of RD programmes: Approaches for assessing the impact of the rural development programmes in the context of multiple intervening factors. Exploring the state of the art and good practice of impact evaluation and measurement', Brussels: EEIG RurEval.
Leeuw, F. and Vaessen, J. 2009
'Impact Evaluations and Development' NONIE Guidance on Impact Evaluation, Washington, DC NONIE –
The Network of Networks on Impact Evaluation; c/o Independent Evaluation Group.
Lipsey, M. W. and Cordray, D. S. 2000
'Evaluation methods for social intervention', Annual Review of Psychology 51: 345-375.
Mayer, S. Hölzl, K. Robubi, A. Enichlmair, C. and J. Marchart 2011
'Evaluierung des Haftungsinstrumentariums des BMWFJ auf Basis des KMU-Förderungsgesetzes',
KMU FORSCHUNG AUSTRIA, Wien.
Pawson, R. and Tilley, N. 1997
Realistic Evaluation, London: Sage Publications.
Peikes, D. N., Moreno, L. and Orzol, S. M. 2008
'Propensity score matching: A note of caution for evaluators of social programs', American Statistician 62(3): 222-231.
Posavac, E. J. and Carey, R. G. 2007
Program Evaluation: Methods and Case Studies, 7th Edition, New Jersey: Pearson/Prentice Hall.
Riley, R., Kirby, S., Meadows, P., Ven, J. v. d. and Barrell, R. 2007
'Evaluation of the macroeconomic impact of Jobcentre Plus and Jobseeker's Allowance New Deals: a feasibility study',
in Department for Work and Pensions (ed) Research Report No 433: A report of research carried out by the National Institute of Economic and Social Research on behalf of the Department for Work and Pensions.
Rosenbaum, P. R. and Rubin, D. B. 1983
'The central role of propensity score in observational studies for causal effects', Biometrica 70: 41–55.
Smith, J. 2000
'A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies',
Schweizerische Zeitschrift für Volkswirtschaft und Statistik 136(3): 1-22.
Tilley, N. 2000
'Realistic Evaluation: An Overview' Presented at the Founding Conference of the Danish Evaluation Society',
September 2000.
World Bank 2006
'Conducting quality impact evaluation under budget, time and data constraints'
Washington: Independent Evaluation Group, The World Bank.

Data sources

aws: monitoring data
Austrian Institute for SME Research: balance sheet database

Authors

Peter Kaufmann
KMU FORSCHUNG AUSTRIA (Austrian Institute for SME Research)
MAIL p.kaufmann@kmuforschung.ac.at

Sabine Mayer
Österreichische Forschungsförderungsgesellschaft FFG (Austrian Research Promotion Agency)
MAIL Sabine.Mayer@ffg.at

Barbara Good and Brigitte Tiefenthaler

Innovation voucher – small is beautiful

Introduction

Technopolis Group Austria had the opportunity to conduct the interim evaluation of the Austrian innovation voucher programme in 2011[1] and to evaluate the Swiss innovation voucher pilot in 2009[2] . In this article, we will present the most important findings of our evaluations, mostly focusing on the programme logic and the programmes' effects.

The Austrian innovation voucher

The innovation voucher programme started in November 2007. The interim evaluation was conducted between the end of 2010 and spring 2011. The methodology used was a mixed method approach: analysis of monitoring data, document analysis, establishing of intervention logic, interviews with stakeholders and SMEs, focus groups with research organisations, participant observation at FFG.

The programme is administered by the Austrian funding agency FFG. A voucher is worth 5.000 EUR, covering 100% of project costs *(no matching funds required by firm)*. Target group: SMEs.

Owners:	Federal Ministry for Transport, Innovation and Technology and Federal Ministry of Economy, Family, and Youth
Budget:	5.2 million EUR p.a. (16.5 million EUR, 2007-2010)
No. of vouchers issued:	903 per annum *(2935 between 11/2007-02/2011)*, no upper limit
Sectoral affiliation:	70% service sector, 30% industry
Firm size:	87% small firms (<50 employees), 13% medium-sized firms *(between 50 and 250 employees)*

The Swiss Innovation voucher

The innovation voucher was launched as part of a business stimulation bill passed by Parliament at the end of 2008 in the wake of the worldwide financial crisis. The pilot scheme was launched in April 2009.

We evaluated the first 30 innovation voucher projects in summer/autumn 2009, shortly after the launch of the programme. The methodology used was a mixed method approach: analysis of monitoring data, document analysis, interviews with SMEs, research organisations and stakeholders, interviews with a control group of SMEs. After the pilot in 2009, the innovation voucher scheme was continued, with a second series of innovation vouchers issued in 2010 and a third series issued in the 2011.

[1] Barbara Good, Brigitte Tiefenthaler, Zwischenevaluierung des Programms Innovationsschecks. Endbericht, 10. August 2011

[2] Barbara Good, Anton Geyer, Evaluation des Pilotprojekts Innovationsscheck. Endbericht an die Förderagentur für Innovation KTI, 16.12.2009

Programme administrated by the Swiss innovation agency CTI. A voucher is worth 7.500 CHF *(approx. 5.000 EUR at the time of launching the programme)*, covering 100% of project costs *(no matching funds required by firm)*. Target group: SMEs.

Owner:	Federal Office of Professional Education and Technology
Budget:	1 million CHF per annum
No. of vouchers issued:	limited to 133 per year
Disciplines:	61% engineering, 23% enabling science (mostly ICT), 11% life sciences, 6% nano and micro technologies (pilot)
Firm size	73% small, 27% medium-sized (pilot)

The programme logic

The programme logic *(often also called intervention logic)* depicts the so-called programme theory, describing the programme's objectives, the anticipated outputs and benefits and the causality linking them. We have established the Austrian programme logic in cooperation with the programme's owners and the funding agency FFG who administers the programme.

The overarching aim of the Austrian innovation voucher programme consists in encouraging SMEs to take up regular R&D and innovation activities, thus enlarging the R&D base of SMEs. From this overarching objective, several more specific objectives have been deduced:

- Stimulating knowledge transfer between SMEs and the science sector
- Closing the knowledge gap between research organisations and SMEs
- Overcoming SMEs' reluctance to get in touch and work with research organisations
- Increasing SMEs' ability and willingness to cooperate with research organisations
- A separate objective was for FFG to attract new clients

The programme's objectives can be found in the first row in *Figure 1*. What is striking is that all objectives have been formulated for SMEs. The research organisations that the SMEs cooperate with and who conduct the innovation voucher project, that is university institutes, non-university research institutes, and universities of applied sciences, do not seem to play a role at all, no objectives having been formulated for them. As a consequence, all the anticipated benefits accrue with SMEs. This of course raises the question: Is there nothing in the programme that might benefit research organisations?

The target group of the Austrian innovation voucher programme are SMEs[3] in Austria *'that do not innovate regularly and do not have their own R&D staff, thus depending on knowledge transfer from research organisations.'*[4] As can be seen from the programme logic, additional characteristics of the target group are that SMEs have not yet cooperated with research organisations and are new to FFG.

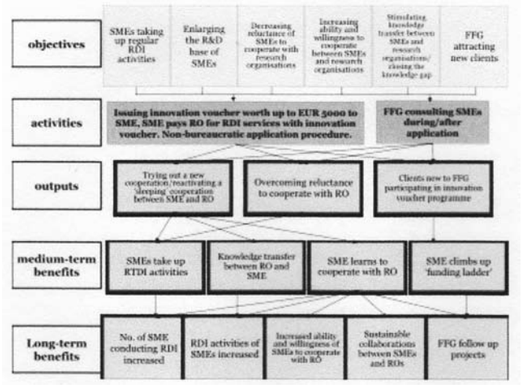
The Swiss innovation voucher, having been modelled on the Austrian instrument, uses the same definition of its target group as the Austrian innovation voucher. However, other than in the Austrian case, research organisations seem to be at least a secondary target group, indicated by the fact that a contract is concluded between the research organisation and the CTI.

[3] The EU definition for SMEs applies

[4] Sonderrichtlinien Innovationsscheck des Bundesministeriums für Verkehr, Innovation und Technologie und des Bundesministeriums für Wirtschaft, Familie und Jugend, Fassung vom 5.11.2010

Figure 1
Programme logic of Austrian innovation voucher programme

Source: Technopolis



The innovation voucher in reality
Target group reached

A very important result of our evaluations concerned the target group actually reached with the innovation voucher – that is small enterprises. In the Austrian case, 86% of innovation vouchers went to firms with fewer than 50 employees. Since FFG does not collect data on the exact number of employees, we had no data on how small the firms actually are. However, there were strong indications that the main group reached were very small enterprises with up to 10 employees, and often up to 5 employees.

The same phenomenon can be observed in Switzerland: Of the innovation vouchers issued in the pilot scheme, 44% went to SMEs with up to 5 employees, 11% to SMEs with up to 10 employees, and 18% of SMEs with up to 25 employees[5].

Clearly, the innovation voucher is an instrument for *very small* SMEs.
Another important result concerned the innovativeness of SMEs that used innovation vouchers. Practically all Austrian SMEs that we interviewed considered themselves innovative. They were all ready to explain why and presented valid arguments. They are innovative in the sense that they are entrepreneurial, looking for business opportunities and trying out new ideas. The research organisations that conduct innovation voucher projects confirmed this result. Practically all Swiss SMEs too thought of themselves as innovative and were ready to explain why. Hence, there is strong empirical evidence that the innovation voucher is not an instrument for not-innovative SMEs.

The typical scenario
The main reason why the innovation voucher is so attractive to very small SMEs is that the innovation voucher is a straightforward – and thus easily accessible – instrument, also to SMEs who have no prior experience with funding or innovation agencies.

In the case of the Austrian innovation voucher, the idea was that an SME would apply for a voucher and receive it if it fulfils a number of criteria. The SME did not need to have a concrete project idea at this stage. When it had developed a project idea or encountered a problem in its business operation, it would take the innovation voucher to a research organisation and contract them to do a project, for which the SME would pay with the voucher. The research organisation would then cash the voucher at FFG.

[5] The figures were similar for the third series of innovation vouchers issued in 2011 see www.kti.admin.ch/projektfoerderung/00029/index.html?lang=de *(We have no figures for 2010)*

Of course, as always, reality is a bit more complex. While the above scenario may have been valid in some cases, in many others it differed. Typically, an SME would h ave an idea or problem in mind and approach a research organisation – often one it already knew, perhaps because the owner had studied there, because it had been recommended by a friend or because there had already been some informal cooperation beforehand. Ideally, the research organisation would also be geographically close. The SME would tell the research organisation about the idea or problem it had, and if happy with each other, the SME and the research organisation would agree that the latter should look into the SME's problem.

At this point, the research organisation – or less frequently the SME – would suggest funding the project with an innovation voucher. The research organisation would write the application, often in cooperation with the SME, which would then submit the application to FFG. After receiving the innovation voucher, the SME would contract the research organisation to conduct the project. The research organisation would handle all the paper work for the SME, which is logical given its experience with applying for public funding. Typically, the SME would invest a fair amount of time in the project, preparing it, and interacting and working with the research organisation[6].

Interestingly, this was the most frequent scenario in the case of the Swiss innovation voucher as well. In both countries the innovation voucher is mainly used to test an idea or a concept. The Austrian programme owners were concerned that many projects would be driven by research organisations in search of funding instead of being driven by SME in need of expertise. Such projects, however, turned out to be very rare.

As a consequence, SMEs find the innovation voucher a very simple instrument while some research organisations complain about the administrative burden it imposes on them. In the case of the Austrian voucher, the burden is exacerbated because FFG communicates with the SME only so all communication goes through the SME. Moreover, the research organisation is responsible for making sure that the innovation voucher project is a 'fundable project as defined by the guidelines[7]', otherwise it will not get its money from FFG after having done the work.

New clients
In Austria, in the period under investigation *(November 2007 – February 2011)*, 83.5% of SMEs who applied for an innovation voucher were new to FFG. New clients are defined as SMEs who have not applied for any FFG funding in the past five years.

The share of SMEs who are new to FFG remains constant over the years – it is regularly over 80%. This implies that the Austrian innovation voucher is far from being a customary right whereby the same SMEs keep applying for innovation vouchers[8]. It is not very surprising that FFG could attract new clients with the innovation voucher because the clientele it has attracted with the innovation voucher - mostly very small SMEs in the service sector – is not the typical clientele of FFG.

The Swiss CTI too could attract new clients with the innovation voucher. In the pilot phase, for 77% of SMEs who had received an innovation voucher it was their first CTI funding. Similarly in 2011, for 71% of SMEs that had applied for an innovation voucher it was their first CTI application.

[6] This was a success factor for good project results
[7] Sonderrichtlinien Innovationsscheck des Bundesministeriums für Verkehr, Innovation und Technologie und des Bundesministeriums für Wirtschaft, Familie und Jugend, Fassung vom 5.11.2010
[8] Austrian SMEs can apply for one innovation voucher per year. In Switzerland, SMEs can only receive an innovation voucher every five years

Follow-up projects

Formal follow-up projects are defined as FFG *(or CTI)* funded projects an SME embarks on after its first innovation voucher project. In the programme logic, formal follow-up projects are regarded as an important effect of the innovation voucher because the innovation voucher aims to encourage SMEs to take up regular innovation and R&D activities, and formal follow-up projects are considered an indication that SMEs have taken up such activities.

There is a temporal and possibly a causal link between the first innovation voucher project and the following FFG *(CTI)* project. In the first case, the positive experience with the innovation voucher may have encouraged the SME to conduct another FFG *(CTI)* project, with the same partner or not. In the second case, the innovation voucher project and the follow-up project are related by content.

In Austria, a total of 15.5% of SMEs who are new to FFG have a formal follow-up project. Not surprisingly, there is temporal effect: SMEs that had their first innovation voucher in 2007 have a formal follow-up project in 27% of cases, and SMEs with their first innovation voucher in 2008 in 24% of cases. In contrast, SMEs that had their first innovation voucher in 2009 so far have a formal follow-up project in only 15% of cases. Clearly, the more time passes, the likelier an SME is to submit another project to FFG.

However, if we look at what types of FFG projects SMEs embark on after their first innovation voucher, we find that it is another innovation voucher in 41% of cases. This is due to the small size of SMEs: Many SMEs are so small that the innovation voucher is the only type of project funded by FFG it can handle. A feasibility study would already be too large for a very small SME. In other words, most FFG projects are not suitable for very small SMEs. Hence, climbing up the 'funding ladder' as envisaged by the programme logic is not possible by many SMEs because they are too small.

On the other hand, the guidelines prevent an SME to use another innovation voucher with the same research organisation the next year. This is, of course, not conducive to building up a sustainable relationship with a research organisation, which is actually one of the aims of the innovation voucher programme. Another consequence is that sometimes an innovation voucher project just stops or is shelved.

Interestingly, in other instances, FFG and ministries are well aware that long-term sustainable relationships between firms and academia need a long time horizon and accordingly are subsidised repeatedly over many years. The consequence is that the smallest firms only get one chance to cooperate with the same research organisation while larger firms are encouraged with large subsidies to form relationships and maintain them. In other words, the smallest firms – those that are supposed to learn to cooperate with research organisations – are subject to the strictest rules. In this instance, the innovation voucher is essentially a break with the funding logic prevailing in the Austrian research and innovation system.

In the Swiss pilot scheme, 57% of SMEs said they had submitted or would submit an application for a regular CTI project[9]. However, since we conducted our evaluation at an early stage of the pilot, we were unable to specify exactly how many applications for regular CTI projects had been submitted and how many had been approved. Indeed, according to information from the CTI, the share of SMEs that conducted a follow-up project was considerably lower in the end – only 13% of the innovation voucher projects had a formal follow-up.

[5] Broadly speaking, the CTI only funds one type of project – regular CTI projects. There is also a feasibility study but it is in essence a smaller CTI project

The Swiss case is perhaps a bit different from the Austrian case because regular CTI projects follow a similar logic as the innovation voucher. Regular CTI projects are joint projects between a firm and a research organisation, with CTI funding going exclusively to the research organisation. They differ from innovation vouchers in that the firm has to cover 50% of the project costs, but the firm's contribution can be in kind. In Switzerland, the innovation voucher is sometimes used to test the feasibility of an idea or concept which is intended to be developed further in a regular CTI project conducted jointly by the same partners.

Having said this, formal, FFG (CTI) funded follow-up projects is too narrow a perspective as an outcome indicator because they do not cover all the activities following the innovation voucher. We heard of follow-up activities financed by the SME itself or funded by other agencies, e.g. at regional level, or contract research carried out by the research organisation on behalf of the SME. Sometimes the cooperation is carried on in an unexpected form, e.g. when a research organisation trained the SME's staff. In Switzerland too, there were follow-up activities not related to the innovation agency, e.g. a SME conducted a follow-up project with another firm and other SMEs had a student follow up the results of the innovation voucher project in a diploma thesis.

Additionality

In Austria, 4 of 20 SMEs *(20%)* would have carried out the project also if they had not received an innovation voucher. In these cases, there is no additionality *(Figure 2)*. We considered this share acceptable, in line with other FFG programmes.

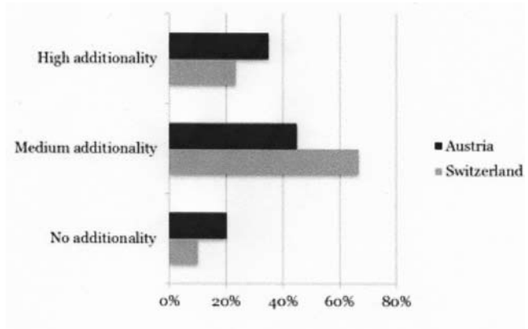
7 of 20 SMEs *(45%)* would not have carried out the project if they had not had an innovation voucher, representing high additionality.

More often, in 9 out of 20 cases *(35%)*, SMEs would have carried out the project without innovation voucher but they would have carried it out differently. In essence, the innovation voucher makes SMEs conduct a larger, better project more quickly at an earlier point in time. Indeed, some SMEs told us that the innovation voucher helped reduce the time-to-market by 50%.

Of course, the innovation voucher is not irreplaceable. SMEs would have found other ways to fund or conduct the project. For instance, one SME would have hired a student over the summer to solve the problem. Others would have tried to integrate the tests carried out in the innovation voucher project in a follow-up project.

Figure 2
Additionality of innovation voucher in %

Source: Technopolis



In Switzerland, additionality was similar *(Figure 2)*. 3 out of 30 SMEs *(10%)* would have carried out the project also without innovation voucher *('no additionality')*, 7 of 30 SMEs *(23%)* would not have carried out the innovation voucher project if they had not received an innovation voucher *('high additionality')*, and 20 of 30 SMEs *(67%)* would have carried out the project differently. Like in Austria, projects would have been conducted later or at a smaller scale. Especially start-ups emphasised that the innovation voucher allowed them to do things in parallel, moving their business faster.
In the Swiss case, we had the opportunity to build a control group. Since, contrary to Austria, innovation vouchers are limited, there was a considerable number of SMEs that had applied for an innovation voucher and would have qualified for one but there were none left. We drew a random sample of 20 SMEs and asked them what they had done after not receiving an innovation voucher. Results showed that the additionality of the innovation voucher is robust.

Other benefits

SMEs

Other effects Austrian SMEs reported were new and intensified contacts with research organisations, follow-up activities outside FFG *(see 0)*, knowledge transfer, triggering innovations and innovation processes in the firm, and less reluctance to cooperate with a research organisation. Swiss SMEs reported similar effects.

Research organisations

Austrian research organisations agreed that the advantages of the innovation voucher laid mostly with SMEs. Nonetheless, they did report some benefits, the most important being follow-up projects and better access to SMEs. In Switzerland too, the main benefits for research organisations are follow-up projects and better access to SMEs in the sense that the innovation voucher 'persuades' SMEs to do a project with the research organisation *(see 0)*.
Both representatives of Austrian and Swiss universities talked of innovation voucher projects as 'a service to Austrian *(Swiss)* SMEs'.

Discussion of the programme logic

The Austrian evaluation showed that the outputs, as conceived in the programme logic, were realised *(Figure 3)*.

- Participation of clients new to FFG in the innovation voucher programme: With more than 80% of SMEs who apply for an innovation voucher, this objective has been achieved.
- Overcoming the reluctance to cooperate with a research organisation: This objective has been achieved although a number of SMEs already had some prior contact with a research organisation.
- Trying out a new cooperation: Trying out a new cooperation was an important motivation for SMEs to get an innovation voucher.

With regard to medium and long-term benefits, there are some problems:

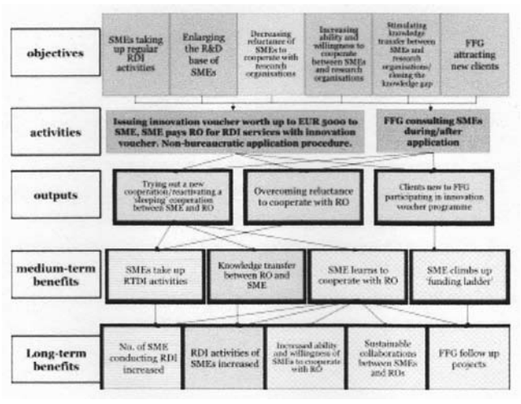
- Most SMEs we interviewed think of themselves as innovative in the sense of entrepreneurial and dynamic. There are strong indications that not innovative SMEs - those not constantly looking for new business opportunities – cannot be reached with the innovation voucher. Hence, the innovation voucher is not suitable for encouraging SMEs to take up innovation and R&D activities. For this reason, we assume that the number of SMEs regularly conducting RDI will not increase. However, the innovation voucher does contribute to SMEs increasing their innovation and R&D activities. This is particularly true for very small SMEs.
- Knowledge transfer between research organisations and SMEs is an important medium-term benefit and a key motivation for SMEs to apply for an innovation voucher.
- SME learns to cooperate with research organisation: A fair number of the SMEs had *(loose)* contacts with research organisations before applying for an innovation voucher project and know exactly how to get what they want.
- Increasing the ability and willingness of SMEs to cooperate with a research organisation. Many SMEs know that it can be very useful to cooperate with a research organisation but have never tried out a formal *(project-based)* cooperation. The innovation voucher is a big incentive to try out such a formal cooperation. In fact, that is why they apply for an innovation voucher.
- Sustainable collaborations between SMEs and research organisations: Typically, contacts between SMEs and research organisations continue but they are often informal and due to the small size of SMEs will not necessarily lead to a formal follow-up project.
- Climbing up the 'funding ladder'/conducting larger and more challenging FFG funded projects: For many SMEs it is not possible to conduct larger FFG funded projects because they are simply too small. They lack the funds and the staff for doing so. For this reason, more than 40% of follow-up projects are innovation vouchers.

Testing the programme logic has shown that not all effects and benefits could be achieved. This is mainly to due the target group reached, which differs from the one anticipated. Nonetheless, the innovation voucher has some very useful benefits for SMEs, mainly helping them to initiate innovation projects and to strengthen relationships with research organisations, so that our overall conclusions about the Austrian innovation voucher programme were largely positive.

Figure 3
Programme logic of Austrian
innovation voucher programme, empirically tested

No pattern: objective/benefit achieved
Cross-wire: objective/benefit partly achieved
Horizontal lines: objective/benefit not achieved

Source: Technopolis



At this point, it will come as no surprise that we had similar results in Switzerland. In particular, we found that in half the cases the innovation voucher had contributed to maintaining, increasing or facilitating RDI activities. This was an important result because maintaining and increasing SMEs' RDI activities was an important motivation for launching the innovation voucher in the first place. Moreover, like Austrian SMEs, Swiss SMEs have heard the message that it is useful to cooperate with research organisations and used the innovation voucher 'to give it a try'. Access to research organisations' expertise was a key motivation for applying for an innovation voucher. As in the Austrian case, our conclusions about the Swiss innovation voucher pilot were largely positive.

Conclusions and recommendations

Our main conclusion in both evaluations was that the innovation voucher is a very useful instrument. Because it is so easy to access, it is particularly attractive for very small firms, which due to their size normally have problems accessing project funding. Very small firms represent a new clientele for both the Austrian FFG and the Swiss CTI. Given the additionality and the effects of the innovation voucher, we recommended continuing the programmes both in the Austrian and the Swiss case. However, we recommended some changes to increase the programmes' effectiveness.

For example, both in Austrian and Switzerland, we recommended limiting the innovation voucher to small SMEs with fewer than 50 employees because its these SMEs that are most in need of the innovation voucher and – the smallest among them – represent a new target group for the agencies. Again in both cases, we recommended neither increasing nor decreasing the value of the innovation voucher. In the Austrian case, we also recommended allowing up to two follow-up innovation vouchers with the same research organisation, in order to increase the chance of establishing a sustainable cooperation.

In both countries, we advised adapting the implementation. For example, in the Austrian case we recommended acknowledging research organisations as a target group. In practical terms, this meant that FFG should stop communicating exclusively with the SMEs and also communicate with the research organisations. Moreover, we suggested that FFG should establish a precise monitoring of firm size, so that more robust analysis of the target group reached would be possible in the future. In the Swiss case, we recommended simplifying the contractual agreements. In Austria, the innovation programme will be continued in all likelihood, with some alterations. In Switzerland, the innovation voucher programme has been continued after the pilot phase. In 2010, the innovation vouchers were limited to the 'clean-tech' industry while in 2011 the innovation voucher programme was open to all SMEs.

With regard to the comparison between the Austrian and the Swiss innovation programmes we made in this article, we could show the large similarities in results. However, this is not particularly surprising because the two economies are structurally similar in terms of firm sizes represented and sectoral mix. More importantly, the Swiss innovation voucher being modelled on the Austrian, the two innovation voucher programmes are very similar, with some differences in implementation. The largest difference between the two programmes lies in the budget, with Austria issuing almost seven times as many innovation vouchers per year as Switzerland, reflecting the countries' different approaches to innovation policy. However, the different budgets do not seem to have influenced the results and there is not really a reason why they should.

Authors

Barbara Good
technopolis |group| Austria
MAIL barbara.good@technopolis-group.com

Brigitte Tiefenthaler
technopolis |group| Austria
MAIL brigitte.tiefenthaler@technopolis-group.com

Anton Geyer and Brigitte Tiefenthaler

Does R&D Project Funding Influence Companies' R&D Location Decisions? Findings from the Evaluation of the Austrian Headquarters Programme

Synopsis

In the evaluation of the Headquarters programme of the Austrian Federal Ministry of Transport, Innovation and Technology *(BMVIT)*, Technopolis took a closer look at the factors that influence companies' decision making about the location of their R&D units. Empirical evidence suggests that direct R&D project funding ranks rather low among companies' internal appraisal factors and hence is rather ineffective in attracting and maintaining R&D headquarters. In order to increase the efficiency and additionality of R&D project funding in the Headquarters programme the evaluators suggest that the programme should address higher-ranking determinants of R&D location decisions in internationally active companies.

History of the Headquarters programme

The Headquarters programme *(HQ programme)* was set up in April 2004 by the Federal Ministry of Transport, Innovation and Technology *(BMVIT)*. The original aim of the programme was to attract company R&D headquarters of international companies to Austria by means of direct R&D project funding. In comparison with the key funding instrument for company R&D in Austria, the FFG Basisprogramm *(BP)*, the Headquarters scheme had more attractive funding conditions. The agency accepted larger than usual project budgets, all of the funding was provided as grant *(i.e. no loan component)*, the funding decision could cover a period of up to three years *(in contrast to the year to year funding in BP projects)*, the financial rules for collaborating with universities and research organisations were more attractive than in the standard BP scheme and – most important - the subsidy was not taken into account for calculating the informal ceiling for BP funding of company R&D per year *(set at 7% of the companies internal R&D expenses)*. Thus for companies that usually would have hit that ceiling, the HQ programme provided a source of real additional R&D aid at a higher marginal aid intensity than they could get in the FFG BP programme.

The HQ programme however did not start as a programme, but as a single firm subsidy in search of a programme. In spring 2003 the late state governor of Carinthia, Jörg Haider, announced that Infineon should receive substantial state aid for their then recently announced decision to transfer their global headquarters of the automotive and industrial electronics business unit to Villach. Villach had already been Infineon's competence centre in this business area for some years and the decision to transfer full responsibility seemed quite logical, a point also raised by the then CEO Ulrich Schumacher at a press conference in Munich on 29 April 2003. Nevertheless, the province of Carinthia and the BMVIT still went ahead and provided state funding for Infineon's move. The path chosen was to grant R&D funds as this was feasible within EU state aid rules.

The ministry wanted to avoid a lengthy and potentially burdensome process to notify the European Commission of direct state aid outside an already notified R&D aid scheme. Therefore the ministry approached the Austrian Industrial Research Fund *(FFF)*, the predecessor of FFG BP, to administer the funding process in line with their already EU notified funding regulations *(FFF-Richtlinien)*. As the total amount of R&D subsidy discussed was rather high and there is a threshold to be taken into account before EU state aid rules might have required an individual aid notification, the subsidy was provided for several, formally independent projects. On 22 April 2004, the FFF decision making body approved a 7.058 m€ R&D subsidy for Infineon's HQ projects. Two days later, the BMVIT also formally mandated the FFF with the administration of the new HQ programme with an annual budget for 2004 of 7.2 m€ *(which at that time had already been fully allocated to Infineon if the 5% FFF management fee is taken into account)*. A second instalment of 4 m€ was granted to Infineon in May 2005 and another 3 m€ in June 2006. For a year and a half the BMVIT HQ programme had exactly one aid recipient.

This precedence of a single political deal begetting an entirely new funding programme – with potentially very attractive funding opportunities for company R&D – subsequently lead to quite some discussions in the ministry and the FFG BP about the development of a coherent framework for applications coming from other companies. Understandably the ministry and FFG wanted to create a level playing field for all. It soon became clear that the number of cases one could expect to actually relocate or set-up new R&D headquarters in Austria might be very limited. Also smaller Austrian companies came into the focus of the programme as the ministry and the FFG wanted to avoid a too narrow orientation of the HQ programme on winning the competition for R&D functions in large, internationally R&D active companies: Why should company size matter? Why should an internationally distributed R&D function matter, when a company plans to substantially increase their R&D capacity in Austria or to expand R&D in new promising technological areas? Therefore, the ministry and FFG broadened the programme's objectives and adapted the HQ rules accordingly.

The main criteria for the assessment of eligible project proposals in the HQ programme were:

- Increase of business R&D expenditure
- Increase of R&D personnel
- Increase of knowledge in new technology
- Increase of knowledge in new application areas
- Knowledge transfer through co-operation
- Sustainability of R&D-activities in Austria

On political levels, the HQ programme always enjoyed strong support, not only from the BMVIT but also from the Austrian Council for Research and Technology Development *(Rat für Forschung und Technologieentwicklung)* who in 2005 stated in their 'Strategy 2010' that they recommend to broaden the scope and increase funding in the FFG HQ programme.

Funding and programme participation 2004 – 2009

Between 2005 and 2009 the FFG was able to allocate about 20 to 25 m€ per annum to HQ projects. In the period 2004 to 2009 a total of 66 companies received 114.8 m€w for 90 projects in the HQ programme. Even in times when other funding schemes came under pressure due to public sector austerity measures following the credit crunch in 2008/2009, the HQ programme did not experience any cuts. On the contrary, the BMVIT and the Federal Ministry of Economy, Family and Youth *(BMWFJ)* topped up the HQ funds by several million Euros, specifically targeting the Austrian automotive industry and meant as a short term measure to bolster the economy.

Did the HQ programme achieve its strategic objectives?

Among the 66 companies funded in the evaluation period there is not a single firm that came to Austria or had established a new R&D headquarters causally in response to the HQ programme. In fact, most beneficiaries *(47)* were Austrian companies. Foreign companies funded in the HQ scheme almost exclusively had a long R&D tradition in Austria. We neither could find any empirical evidence that the HQ funding actually triggered or contributed to a significant increase of R&D activity in the companies nor that it was causal for setting up the R&D activities in new technological areas. The companies' strategic decisions to go into new R&D areas or to substantially increase R&D activities predated the application for funding and were well under way *(or even already fully in place)* when the companies received their HQ funding. The „success'of the programme hence mainly reflects the specific selection criteria in the HQ programme, funding projects of companies that recently made significant additional investments in R&D. To put it in other words: The increase of companies' R&D activities triggered public R&D funding but not the other way round. We could find only very limited evidence for real knock on effects of HQ funding on business R&D. The main effects of the HQ funding that we could identify were to reduce the economic risk of the R&D investments and – in certain cases, but by far not in all – to speed up the project related R&D activities in the companies. These impacts are broadly in line with the impacts expected from projects funded in the standard FFG BP scheme. We concluded that from an efficiency perspective of public R&D funding, the HQ programme could not prove any additional impacts compared to BP funding. In consequence we could not see any empirically based justification for applying the more generous HQ funding principles over the BP scheme. At the end of the day HQ funding was barely more than some sort of progressive gratification for companies that had increased their R&D activities.

In this respect we would like to stress that realistically one also shouldn't *(and couldn't)* expect more than that which was achieved in the HQ funded projects: Contrary to conventional wisdom that seems prevalent in much of the R&D policy community today, we found that *(the prospect of)* direct R&D project funding does not rank high at all on the list of considerations when companies take strategic decisions about R&D locations or when they substantially increase R&D capacity at existing R&D locations. The firm's quest for public R&D project funding only becomes dominant later in the tactical domain when companies aim to optimise the financing of R&D activities on operational levels.

Critical factors that influence R&D location decision in companies

If availability of public R&D project funding is not a critical strategic issue in R&D location and expansion decisions by companies, what factors are? As part of the evaluation we spoke to the representatives of 29 companies in the HQ programme of which 27 companies had actually received HQ funding. These companies accounted for 64% of all HQ funding in the programme period 2004 to 2009. Taking the HQ project as a starting point, the interviews covered the business and R&D background, the history and the decision processes that led to the establishment or the expansion of the HQ R&D area in the company. From the analysis of the interviews we derived the following key determinants for decisions about R&D locations and the substantial expansion of R&D activities:

1. Legacy: there is only one relevant R&D location (and no need for another one)

A large share of companies we spoke to about their HQ projects have only one relevant R&D location. They are also

quite happy with this situation. Even companies with various international production sites, very high export quota and key clients all over the world mentioned in the interviews that there was no need for a second research location at all as the anticipated coordination and knowledge transfer costs between R&D locations would far outstrip any potential economic gain. Several Austrian based companies with more than 1500 staff at home locations and several hundreds staff abroad said that they considered themselves being „too small'for more than one R&D location. This finding is very much in line with evidence from many large international companies over the past decade which have significantly reduced the total number of R&D locations in order to streamline their internal R&D management processes and reduce R&D exploitation costs.

2. Competences: drawing on infrastructure and know-how already available at existing R&D locations
Not only does legacy count when there is only one R&D location but company history also matters when there are several R&D sites. The location of any expansion of R&D activities in a certain technological area or any newly established R&D unit is strongly influenced by the infrastructure and the competences available at already existing R&D sites. Companies attach new R&D units to locations where there is maximum in-house know-how already available. The most generous R&D tax breaks and subsidy schemes do not make sense from a company perspective if the company doesn't have the R&D competences available and ready on the location in question. Available *(knowledge)* infrastructure does also count significantly: Some companies reported that in the past they had already relocated R&D and engineering functions from abroad back to Austria since the competences und *(human and infrastructural)* resources available abroad could not live up to the companies' expectations on quality.

New R&D locations mainly emerge as a result of company take-overs. Whether the parent company then keeps the newly acquired R&D location or either shuts down, sells or aims to transfer the R&D unit to Austria mainly depends on the specific capabilities and competences of the respective new and the existing R&D locations: Is there any additional value in keeping the acquired R&D location? What are the prospects of transferring the respective know-how to Austria or to another company R&D location within the company? In several cases we could find evidence that companies had consolidated smaller, sub-critical R&D units in stronger R&D locations.

3. Proximity: creating synergies through concentration of R&D and production
Another important determinant for R&D location decisions is proximity of research and production. Especially process innovation requires a direct link between R&D and operations. Hence, companies in process industries often seek synergies by locating their *(new)* R&D units close to the main production sites. In many cases this is often also the firms' historical headquarters.

4. Markets: seeking opportunities and reacting to pressures
Companies not only go abroad in order to take advantage of cheaper investments or labour but also – perhaps predominantly – because they expect new business opportunities and a promising pathway to long term growth. Especially the ongoing trend to open company representations and production sites in south-east Asia can rather be explained by the companies' expectations of potentially huge and fast growing industry and consumer markets than merely by cost advantages sought in the short term.

In consequence, cost factors alone do not drive Austrian companies abroad as far as their R&D functions are concerned. Several companies reported that they take advantage of cheaper labour costs in Eastern Europe for some of their less sophisticated development and engineering jobs. Critical R&D, however, remains in Austria, as the companies want to keep their cards close to the chest in all areas of strategic long-term importance.

Companies do consider moving some of their R&D units abroad if additional pressure emerges. This can be, for example, in the form of a majority foreign shareholder or foreign owner who has an interest to strengthen the company's visibility and clout on their home market abroad.

5. Human resources: access to qualified R&D staff and technical personnel
Many of our interviewees stressed the critical issue of highly qualified R&D staff and technical support personnel in company appraisal processes of R&D location decisions. Companies seek proximity to technical universities and universities of applied sciences; they consider locations quite disadvantaged if they are remote to teaching and research establishments or unattractive for young university recruits *(e.g. rural areas)*. Some companies mentioned having already faced severe difficulties in attracting university graduates to their headquarters location in peripheral Austrian regions.

6. Ownership: preferences of owners and board members
Not always are economic decisions taken purely based on fully rational considerations. Or to put it more precisely: there is considerable room for discussion on what accounts for rational economic behaviour. What one person might see as the consequent pursuit of a bold strategic vision another person might understand as a pet project of owners or members of the board. R&D involves risks. For example, we were able to identify some HQ companies whose R&D activities were clearly more a bet on the future rather than staying on the beaten track. The important point we want to raise in this respect is: owners' and senior management's preferences do count when it comes to R&D related decision making.

Especially in family owned companies, the owners and the owner families are often very much attached to the location where the company originated and their historic headquarters. New R&D facilities are then often established in close proximity or even on site of already existing company facilities. In such cases, the question why the company did not look for a more suitable research location abroad or somewhere else in Austria was met with quite some incomprehension: Why? The owner/the owner family is strongly committed to their traditional location. Why should they want to move key business functions?

There was also one case in our HQ company sample, though, where the owner is quite mobile himself. In this company there is an ongoing discussion to move parts of the R&D activities to a location in South-East Asia as the owner lives for some part of the year there anyway.

Some interviewees even mentioned that the nationality and the professional background of the board members seem to influence location decisions, especially if they have had previous operational experience and therefore a good knowledge of the Austrian production and research capabilities.

7. Costs and tax incentives
All the factors mentioned above bear influence on major R&D *(location)* decisions and they are mainly related to perceived opportunities, not perceived costs. From a business perspective, this is completely rational. Businesses strive

for growth and pursue opportunities in order to increase profits and profitability. Reducing costs only becomes an issue after key strategic decisions have been taken and only as far as costs or financial incentives can be directly factored into the appraisal of projects or locations. Since project related R&D subsidies usually can not be taken for granted, other factors, such as the overall effective business tax rate or indirect R&D funding via tax breaks, are more relevant than R&D project funding in the companies decision making processes. Among our interviewees there was at least one, whose company specifically set up operations in Austria, because of a more advantageous business profit tax regime in Austria compared to the owner's country of origin. Several other interviewees mentioned that the board considers attractive tax regulations rather more critical than R&D project subsidies. However, R&D subsidies play a crucial role in securing the internal financing of specific projects. Direct R&D state funding might not change in every case the overall spending on R&D in the company but certainly the ranking of projects within their internal R&D project portfolio.

Policy implications

The most relevant message from our findings in the evaluation of the HQ programme that we would like send to R&D policy makers is: Don't be scared! Companies' R&D location decisions need not keep you awake at night! The prospect of direct R&D project funding will most certainly not tip the scales in one direction or the other. However, there is indeed an important role to play for you in strengthening and maintaining a sustainable framework that keeps Austria attractive as a powerful R&D location for internationally active companies. Just attach strings to direct R&D subsidies that get companies more effectively hooked on Austria as an R&D location.

As a specific recommendation we suggested that more favourable funding conditions than those available in the FFG BP scheme should be subject to structural requirements such as the establishment of long-term research co-operations with Austrian universities and research institutes. In the long run, this would clearly create spill-over effects beyond the funded companies, as *(young)* researchers will be trained in science-industry co-operations. In turn, the academic and research sector institutions become more attractive for other potential industrial partners. One needs to keep in mind, though, that in Austria there are already plenty of funding opportunities available for strengthening science-industry relations and any new programme should be subject to a thorough ex-ante programme evaluation.

For internationally active companies that aim to increase their R&D activities without any stronger linkages to the overall Austrian innovation system, our findings suggest that – considering additionality and efficiency of public funding – the already available funding opportunities in the FFG BP scheme provide sufficiently attractive incentives.

Authors

Anton Geyer
technopolis |group| Austria
MAIL anton.geyer@technopolis-group.com

Brigitte Tiefenthaler
technopolis |group| Austria
MAIL brigitte.tiefenthaler@technopolis-group.com

Ilse Marschalek, Katharina Handler
and Magdalena Strasser

Communicating nanotechnology to European youth – evaluation results of the NANOYOU project

Abstract

The FP 7 funded project NANOYOU had to find ways to communicate nanotechnologies to young people and develop adequate materials for doing so. The Centre for Social Innovation *(ZSI)*, as one of the project partners, conducted a baseline survey in order to gain insight into what European youths in general know about nanotechnologies and their opinion and attitude towards these technologies. We also evaluated the two stages of outreach activities within the project to find out whether young people's knowledge increased as a result of the project and whether their attitudes concerning nanotechnologies had changed.

This article displays considerations about the methodological approach of evaluating knowledge and attitudes in a field which is widely unknown by the respondents. Furthermore, it summarises results of the European baseline survey, carried out at the beginning of the project compared with evaluation results to show the project's impact on knowledge and attitudes of young people.

Introduction

Several studies and technology assessment processes revealed similar results: People have a very limited knowledge of nanotechnologies *(Vierboom et al., 2008)*. The same seems to be true for young people *(Zöller, 2008)*. Some youths have heard the word 'nano' before and some of them associate it with something very small, but only very few young people really understood its dimension. When asked, accordingly, people find it rather difficult to assess the implications of nanotechnology. Young EU citizens found it difficult to assess the balance of risks and benefits in the case of innovations in the field of nanotechnology. However, the proportion of respondents who think that there are more benefits than risks is significantly larger than the proportion of people who think the opposite *(Eurobarometer Flash Survey, 2009, p. 34)*. So it can be concluded that knowledge and interest are limited and no negative expectations are prevalent.

However, we found that awareness and understanding of the science and technology behind new technologies is relatively unimportant for public perception and opinion-forming. The possible impact of a technology on their own life or that of close relatives and friends is of a much higher importance *(nanoBio-RAISE, n.d.)*.

In general 'the tendency is to overemphasise risks when benefits are not clear and underestimate or accept risks if the product is available on the market and significant personal benefit is experienced from its use'. *(Bucchi & Trench,*

2008) The challenge is to provide sufficient information outweighing possible negative and positive impacts of certain products and to link the discussion of these applications to the daily life of young people.

Communicating nanotechnologies to youth

For communicating a future technology such as 'nano' which is, as often stated, only in its infancy, the future generation is considered as the central public. Young people are addressed mainly because of two reasons. Firstly, they are viewed as a critical public who either accepts or rejects new technologies and who will accordingly be future supporters and consumers or not. And secondly, they are considered as the future engineers and scientists and thus, communication activities have to spark their interest for them to become engaged in the various sectors and fields of nanotechnologies.

ELSA

In line with current developments regarding concepts of public understanding of science *(Bucchi & Trench, 2008)*, within science communication and technology assessment the public should be involved in activities rather than just be informed about the matter in question. This involves getting into a dialogue with the public and also giving room for discussing possible societal impacts. An expected consequence of this approach and the main goal of science communication is a better understanding of certain fields of science. A number of ethical, legal and societal aspects *(ELSA)* are often raised with regard to nanotechnologies. This includes the issue of weighing potential benefits against potential costs, the distribution of benefits and costs among the population, concerns about personal freedom, the environmental impact, the control over the development of nanotechnologies, and many others. Communication activities therefore have to provide balanced information which should neither be too enthusiastic nor too scaring *(Marschalek, Moser, & Handler, 2010)* but at the same time should raise awareness of the complexity of ethical, legal and societal issues associated with nanotechnologies among the target group youth.

NANOYOU

The FP 7 funded project NANOYOU[1] had to find ways to communicate nanotechnologies to young people and develop materials to support this endeavour. The aim of the project was to inform them about nanotechnologies and their applications and also invite them into a dialogue about these evolving technologies. The youths should become aware of opportunities but also of societal impacts and possible risks or uncertainties. Therefore, in addition to gaining knowledge about nanotechnologies, discussing ELSA – ethical, legal, social aspects of nanotechnologies – was an explicit project goal.

The project was designed to take place in two stages of outreach activities, each involving 20 pilot schools across Europe. The majority of participating pupils were between 14 and 19 years old, some schools participated with lower grade students. The project team had to develop adequate materials and instruments best suitable for reaching young people and teaching them about nanotechnologies. We, as one of the project partners, conducted a baseline survey in order to find out what European youth in general know about nanotechnologies and what their opinions and attitudes about this topic are. The outcomes of the baseline survey *(Bonazzi, 2010)* supported the project team for the development of the outreach materials. Those materials contained various kinds of tools *(all available on the project's portal[2])*, such as an introductory video, sets of games *(e.g. role playing game or jigsaw puzzle)*, guides for hands on experiments, posters, PowerPoint presentations, as well as virtual tools *(e.g. virtual dilemmas or virtual experiments).*

[1] NANOYOU Communicating Nanotechnology to European Youth, CSA-SA 233433

[2] www.nanoyou.eu

Before each outreach phase, we invited the teachers of the participating pilot schools to a two day teacher training to learn more about the materials and be prepared for its future usage in the classroom. In addition, we seized this opportunity to assess the material in terms of practical and pedagogical applicability. We used pre- and post questionnaires to evaluate these training activities.

We also evaluated the two stages of outreach activities to assess the materials used in the project and to find out whether young people had a more extensive knowledge on the topic after the outreach activities and whether their attitudes about nanotechnologies had changed.

We faced two great challenges: Identifying changes in knowledge about a field, which is in general hardly known at all. Moreover, and even more difficult: Finding out about young people's opinions and attitudes without influencing them? Science communicators have to inform but at the same time leave room for dialogue and criticism. Therefore, in developing outreach materials a priority was to provide students with balanced information. In terms of evaluating young people's attitudes, we had to find ways to learn more about their knowledge and how they assess the information they have available.

1. Methodology

a. Baseline survey

Before developing evaluation instruments, we carried out an extensive literature research, conducted focus group discussions with young people and interviews with experts from various related fields. For the focus group discussions and the further development of materials that should enhance arguments we had to find a simple way to provide information that makes basic principles of nanotechnologies comprehensible and also stimulates critical thinking on ELSA. In consultancy with scientific project partners we worked out a definition of nanotechnologies, illustrating size and properties, selected a range of examples of nano products and summarised basic information on main fields of application. In the focus group discussions the group was confronted with these materials and was asked related questions. For example, they were shown five pictures including short descriptions of application examples and questions were raised, such as the following:

What do you think about it?
Is it rather a good thing or a bad thing?
Do you see a benefit or threat?
Do you think it outweighs chance or risk?
Do you understand what the nanotechnologies component is?
Would you personally use it?
Why yes, why not?
Do you think that the product already exists?

The examples that worked well were also integrated into the online questionnaire.

To give an insight into tendencies of their attitudes, respondents were also asked to move a slider according to their answer to questions like the following:

What do you think? The jacket is rather
'A good thing' or 'A bad thing'
'Has more benefits' or 'Has more risks'
'I'd like to have it' or 'I wouldn't want to have it'

In addition to questions about attitudes on technologies and nanotechnologies in general, young people's knowledge about nanotechnology was evaluated in form of a quiz. The correct solutions including explanations were given at the end of the questionnaire. We used questions like:

What is a bucky ball?

- A nanoparticle made of gold
- ***A carbon molecule C60***
- A ball used for a tennis-like game in India
- Another word for atom nucleus

Buckminsterfullerene is a hollow sphere of 0.7 in diameter formed by 60 carbon atoms formed like an old styled football. It is an artificial nanoparticle with specific properties that can be used for different applications like medicine.

A little robot that can build a material atom by atom…?

- ***Is fictional, it does not exist***
- Exist in experimental form in labs
- Is used in high tech medicine and drugs
- Is used by NASA in space
- Don't know

In Eric Drexler's book 'engine of creation' *(1986)* he first used the words 'nanobots' and 'assembler', autonomous machines on nanoscale. Such machines are still science fiction.

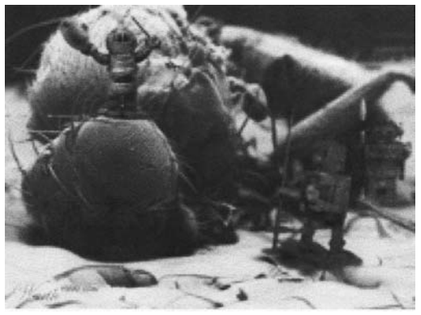


Figure 1
Quiz questions and solutions

The questionnaire was open and not restricted to pre-defined participants. The main idea was to reach young people who had not undergone teaching or communicating activities on nanotechnologies before. Results of this survey were considered in the development of instruments and materials to be used in the two stages of outreach activities at pilot schools.

b. Evaluation activities and instruments

The evaluation instruments of the two outreach stages of the project were manifold. They included teacher interviews, pre- and post evaluation of teacher trainings, data of teachers' documentation, school observations etc *(see overview table below)*. These activities were predominantly used to assess instruments and formats carried out within the project and individual experiences with the activities. In this article results are not reported in detail. The main focus was on assessing young people's knowledge and attitudes on nanotechnologies throughout the project activities. The baseline survey built the basis for comparison. It was completed by participants who had not explicitly experienced any outreach activities before *(see point a)*. For knowledge assessment we used the questions from the quiz again and were therefore able to compare these results with the outcomes after the outreach activities and detect whether pilot school students were able to answer more questions after the project activities. Questions also contained their estimations and assessments of the effects of nanotechnologies. Results of the baseline survey of non-pilot schools and of the evaluation survey with pilot schools were compared. In order to be able to compare results of the different surveys we looked therein at the four most present countries and compared knowledge of students before and after NANOYOU outreach activities in Austria, Romania, Italy and Spain, because of the high numbers of respondents from these countries.

We also conducted focus groups with pilot school participants. To find out what they had learned we asked students to formulate their definitions of nanotechnology and afterwards let them decide for the most appropriate one. We discussed examples and pros and cons of nano products and other aspects related to nanotechnologies and the young people's awareness and knowledge regarding ELSA.

The table below shows an overview of evaluation instruments: ***Table 1*** – evaluation activities overview

Sources of data		
Survey instruments	**respondents**	**countries**
Pre-online questionnaire (WP 1)	1969	in more than 25 countries
Post-online questionnaire (post-inquiry stage one)	486	All 15 pilot school countries of stage 1
Post-online questionnaire (post-inquiry stage two)	457	All 20 pilot school countries of stage 2
Paper-and pen pre and post questionnaires during teachers training days	46	All 20 pilot school countries of stage 1 and 2
Documentation by teachers	23	All 15 pilot school countries of stage 1
Qualitative data (in form of written transcripts as well as in written interpretations for reports D 7.3 and 7.4.)		
Numbers		**countries**
Semi-structured interviews with teachers	29	All 15 pilot school countries of stage 1 + Hungary, France, Finland, Czech Republic (new participating pilot school countries in stage 2
Focus group discussions with students	Six groups (about 50 students from 12-18 years)	Italy, Romania, Spain, Austria, UK
Message protocols on basecamp (internal webportal) - with feedback to trainings activities	1 protocol	all = 20
Observation templates (composed during site visits) including project products, like posters, presentations, role cards etc.	7 classes (11-18 years)	Italy, Romania, Spain, Austria, UK, Hungary, Finland

2. Evaluation results

a. Knowledge

In general, results of the baseline survey confirmed results of former studies. *(Grobe, 2007)*; *(Hanssen, Walhout, & van EEST, 2008)*. It showed that most young people have at least heard about 'nano' but they often do not have a deeper understanding of the topic. Their knowledge about nanotechnologies is predominantly related to specific NT products. They can name different products and applications but cannot give explanations of the technology behind.

In answering the quiz the participants of the baseline survey gained average results, for example 53% of the older female group could answer at least four questions correctly. Young boys tended to know a little more, but it turned out that young boys often guessed correct answers in the quiz rather than indicating 'I don't know'.

However, we could see that young people's interest in nanotechnologies is bigger than their knowledge and they would like to know more about it in the future. Reaching out for young people with the topic of NT is related to interest, either by teachers or youths themselves.

Evaluation results after outreach activities

In comparison with the baseline survey, results after the NANOYOU outreach activities show differences in knowledge and attitudes. Including all participating countries, we compared numbers of correct answers to the online nanotechnology quiz of the baseline results *(2009)* with answers to the same quiz after experiencing NANOYOU outreach activities stages 1 and 2 *(2010 and 2011)*.

If we look, for example, at the answers of males between 15 and 25 *(or 23 in the sample of 2011)* we see the following numbers. **Table 2** – Comparison of Percentage of correct quiz answers *(male participants 15-25 in 2009 and 2010 and 15-23 in 2011)*

Number of correct answersproblems	Baseline results Males 2009 15-25		After stage 1 Males 2010 15-25		After stage 2 Males 2011 15-23	
0	1	34%	4	21%	0	14%
1	5		1		1	
2	12		5		5	
3	16		11		8	
4	29	66%	19	79%	18	86%
5	23		26		36	
6	14		34		32	
	100%		100%		100%	
Persons	478		209		201	

The results in *table 2* show that while in the baseline survey in 2009 only 66% of males between the ages of 15 and 25 could answer four or more of the six questions correctly, this number increased to 79% in 2010 after the first stage of outreach activities was completed and further increased to 86% of males between the ages of 15 and 23 in 2011 after completion of the second stage of outreach activities.

In both age and gender groups students in 2010 can answer more of the same questions correctly.

Example for Knowledge comparison by gender and country Austria:
Table 3 *– Austria- Percentage of correct answers to single quiz-questions sorted by survey and gender Source:* NANOYOU students' questionnaires 2009 + 2010

Percentage of correct answers to:	WP 1 Females (pre 2009)	WP 7 Females (post 2010)	WP 1 Males (pre 2009)	WP 7 Males (pre 2010)
What is bigger – an atom or a bacterium?	83	91	88	89
What can you use to see nanoparticles?	81	97	91	100
What is as small as one nanometre?	50	61	45	67
Do nanoparticles occur naturally?	73	88	74	89
What environment is needed to manufacture computer chips?	57	70	84	75
The properties of a material depend…	67	61	51	67

Results show that Austrian students who received information and participated in outreach activities gained a basic knowledge and know more than their peers who did not experience NANOYOU outreach activities. More of them can give correct answers to simple questions. For more complex questions no increase of knowledge is evident.

Main results

The NANOYOU outreach activities have an impact on the knowledge of students. In general, students achieved rather good results in the knowledge quiz after NANOYOU teaching activities. Gender differences blur with age: Older Girls and boys could nearly equally answer the quiz questions correctly. In stage 1 of outreach the boys of the younger age group tended to know more than their female peers, in stage 2 it was the other way around, namely younger girls tended to know more than boys.

When comparing the knowledge of students who did not experience NANOYOU outreach activities and the knowledge of students who experienced NANOYOU outreach activities tendencies in increased awareness and knowledge become visible. Students noticed a difference in their knowledge concerning nanotechnologies before and after taking part in NANOYOU activities.

Even those students who do not believe their knowledge to be more extensive than before estimated a raised awareness concerning nanotechnologies and their implications. Most of them also had an increased interest in nanotechnologies and stated that they would look for further information on this topic and pay more attention to issues raised in this area.

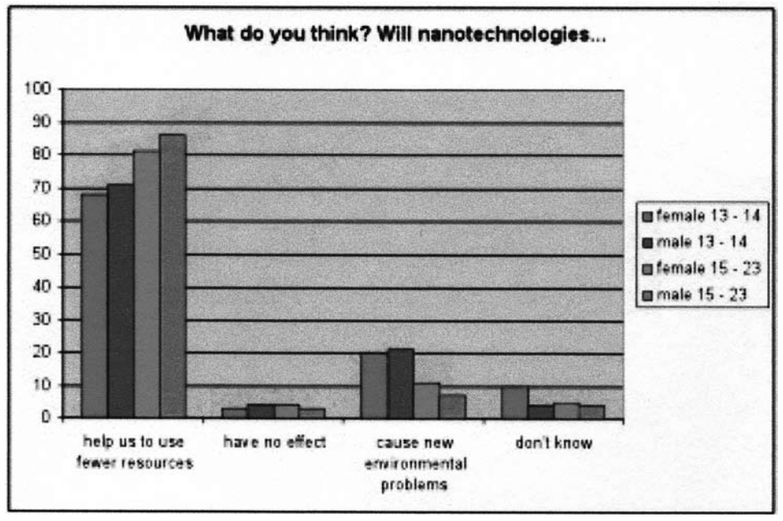
b. Attitudes

In the baseline survey and in the outreach evaluation we asked for associations, attitudes and opinions of young people on nanotechnologies and compared the results.

In the following, some questions and the according results from the online survey are presented:

- To the question:
  'Will nanotechnologies help us to use fewer resources or have no effects or cause new environmental problems?' students gave the following answers:

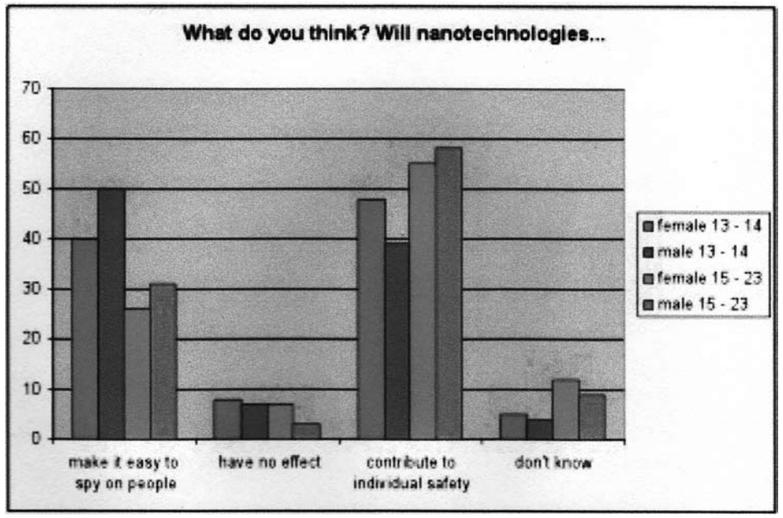
Figure 2 – Attitudes concerning nanotechnologies split in gender and age, questionnaire NANOYOU 2011, in percentage



The majority of students think that nanotechnologies will help us to use fewer resources *(81%).* More than 10% think that new environmental problems will be caused by nanotechnologies.

- To the question:
  'Will nanotechnologies make it easy to spy on people or have no effect or contribute to individual safety?' students gave the following answers:

Figure 3 – Attitudes concerning nanotechnologies split in gender and age, questionnaire NANOYOU 2011, in percentage



31% of the students think that nanotechnologies make it easy to spy on people, although 54% think that they will contribute to individual safety. Younger students, especially male, show more concerns.

Main results

In general, students became aware of ethical, legal and social aspects *(ELSA)* although not all of them feel sufficiently informed.

According to the online questionnaire, in principle, students have rather positive attitudes concerning nanotechnologies, but they are aware of some negative impact that could affect our lives or that a positive impact will not affect all individuals or countries.

Younger students are more critical and have more concerns compared to the older age groups.

NANOYOU activities seem to support a positive self estimation of the students about knowledge on ELSA. After taking part in NANOYOU activities students themselves estimate their knowledge and understanding of ELSA as sufficient.

There is a tendency among the students to support the further development of nanotechnologies but not without limitations and only for the more important purposes.

After the outreach activities students' developed their awareness about nanotechnologies. No explicitly negative attitudes evolved in the outreach activities. Students in general show many concerns about ELSA and were able to consider pros and cons.

However, ELSA discussions should not be dealt with in an isolated way. ELSA discussions on nanotechnologies could be either embedded in wider discussions on ethics and contribute to awareness-raising on ethical issues in general or education in ethics could build the basics for nanotechnology debates on ELSA.

3. Conclusion

The evaluation activities within the NANOYOU project included a variety of instruments, integrating different related groups of respondents and different stages of activities. We developed instruments that on the one hand provided information and on the other hand stimulated discussion to gather young people's opinions on nanotechnologies and their implications. We brought together test results of specific quiz questions with self estimations on knowledge gain and awareness.

We were able to show differences in knowledge and give insight into attitudes and awareness of young people on nanotechnologies after experiencing communication and outreach activities. We were also able to show that results concerning their general attitude towards nanotechnologies and their advantages and disadvantages did not change to the one or the other extreme. Young people did not evaluate nanotechnologies in a much more negative or positive way than before the project, although their knowledge and awareness had changed.

One issue the project could not sufficiently deal with was how to approach the younger age groups. Results of the baseline survey already showed that 10 to 14 years old youths are difficult to reach. During the pilot phase, although lower grades participated, nanotechnologies could not be discussed accordingly. It is not only difficult to integrate nanotechnologies in their school curricula, also most young people in this age group do not yet have the basic terminology and background understanding of science in general.

However, results clearly confirm previous studies *(Satterfield, Kandlikar, Beaudrie, Conti, & Harthorn, 2009)* showing that more information does not necessarily lead to more scepticism or more acceptance. Already young people are able to give differentiated opinions when provided with balanced information. Practical examples related to their daily life supported the discussion, made nanotechnologies easier to understand but at the same time allowed for reflection on possible impacts and ELSA. Still, selected examples only relate to a certain topics. Each example only covers certain aspects. Therefore the examples have to be chosen according to societal dilemmas that should be discussed.

Finally it could be shown that an interdisciplinary development of outreach activities and evaluation instruments tailored to the respective target groups helps to understand young people's opinions and attitudes, even in a field which is still widely unknown to them.

4. Acknowledgement

We would like to thank all NANOYOU partners for their support and all participating teachers and students for their collaborations.

References

Bonazzi, M. (2010)
Knowledge, Attitudes and Opinions on Nanotechnology across European Youth –
Analysis from a specific survey carried out in 25 EU countries. Brussels: EC/DG Research Directorate G –
Industrial Technologies. Unit G.4 – Nano-and converging Sciences and Technologies.
Bucchi, M., & Trench, B. (Eds.). (2008)
Handbook of Public Communication of Science and Technology. Routledge.
Eurobarometer Flash Survey. (2009)
Young people and science. Eurobarometer Flash Survey 239.
Retrieved from ec.europa.eu/public_opinion/archives/flash_arch_en.htm#239
Grobe, A. (2007)
Die Zukunft der Nanotechnologie – Erwartungen von Konsumenten und Experten.
Apunto, 12/1 2007/08Dr Antje Grobe, Leiterin Nanotechnologie, Stiftung Risiko-Dialog, 16–17.
Hanssen, L., Walhout, & van EEST, R. (2008)
Ten lessons for nanodialogue. Rathenau Institute.
Retrieved from www.rathenau.nl/uploads/tx_tferathenau/Ten_lessons_for_a_nanodialogue_2008_01.pdf
Marschalek, ilse, Moser, P., & Handler, K. (2010)
Between euphoria and fear: Balanced communication of nanotechnologies to youth.
Presented at the SciCom2010. Möglichkeiten und Grenzen der Wissenschaftskommunikation.
Retrieved from www.scicom.co.at/2010/cms/website.php?id=panels/panel4.php
nanoBio-RAISE. (n.d.)
Public Perceptions and Communication about Nanobiotechnology.
Retrieved from files.nanobio-raise.org/Downloads/NanoPublicFINAL.pdf
Satterfield, T., Kandlikar, M., Beaudrie, C. E. H., Conti, J., & Harthorn, B. H. (2009)
Anticipating the perceived risk of nanotechnologies. Nat Nano, 4(12), 883. doi:10.1038/nnano.2009.369
Vierboom, C., Härlen, I., Simons, J., Zimmer, R., Hertel, R., & Böl, G.-F. (2008)
Wahrnehmung der Nanotechnologie in der Bevölkerung. Berlin: Bundesinstitut für Risikobewertung.
Zöller, K. (2008)
Jugendforen Nanomedizin. Chancen und Risiken, ethische und soziale Fragen der Nanomedizin
aus der Sicht junger Erwachsener.

Authors

Ilse Marschalek
Centre for Social Innovation
MAIL marschalek@zsi.at

Katharina Handler
Centre for Social Innovation
MAIL handler@zsi.at

Magdalena Strasser
Centre for Social Innovation
MAIL strasser@zsi.at

IMPRESSUM
Herausgabe und Versand
Dr. Rupert Pichler, Dr. Sonja Sheikh, Mag. Klaus Zinöcker, Mag. Leonhard Jörg
Plattform Forschungs– und Technologieevaluierung, Schlickgasse 3/12, A-1090 Wien, office@fteval.at
Gestaltung
to-pixelate, carmen@to-pixelate.org
Druck
Remaprint, Neulerchenfelderstrasse 35, 1160 Wien

FÜR DEN INHALT DIESER AUSGABE VERANTWORTLICH
Plattform Forschungs– und Technologieevaluierung: Mag. Klaus Zinöcker

fteval

PLATTFORM FORSCHUNGS– UND TECHNOLOGIEEVALUIERUNG

Die Plattform Forschungs- und Technologieevaluierung ist eine Initiative der folgenden Organisationen:
Österreichisches Bundesministerium für Wissenschaft und Forschung (bmwf), Bundesministerium für Verkehr, Innovation und Technologie (bm:vit), Bundesministerium für Wirtschaft, Familie und Jugend (bmwfj), Austrian Institute of Technology GmbH (AIT), AQA – Österreichische Qualitätssicherungsagentur, Austria Wirtschaftsservice (aws), Christian Doppler Gesellschaft (CDG), Österreichische Forschungsförderungsgesellschaft mbH (FFG), Fonds zur Förderung der wissenschaftlichen Forschung (FWF), Joanneum Research,KMU Forschung Austria, Ludwig Boltzmann Gesellschaft, Rat für Forschung und Technologieentwicklung, Technopolis Austria GmbH, Österreichisches Institut für Wirtschaftsforschung (WIFO), Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) Zentrum für Innovation und Technologie GmbH (ZIT), Zentrum für soziale Innovation (ZSI).

Im Rahmen der Plattform werden Themenstellungen zur Forschungs– und Technologieevaluierung erarbeitet und – z.T. unter Einbeziehung namhafter ExpertInnen – in einem Fachkreis diskutiert. Der Newsletter beinhaltet Fachbeiträge zu Fragen der forschungs- und technologiepolitischen Evaluierung. Die Herausgabe erfolgt in zeitlicher als auch inhaltlicher Abstimmung mit Plattform-Veranstaltungen, um die Synergiewirkungen eines breiten Austauschforums zu nutzen.

Plattform Forschungs– und Technologieevaluierung
Schlickgasse 3/12, A-1090 Wien

www.fteval.at