BACKGROUND PAPER

ESF performance target setting and adjusting in social inclusion

Client
European Commission
DG Employment, Social Affairs and Inclusion
Unit Impact Assessment, Evaluation

Date
16 May 2013

Author
Drs. Wim Zwinkels
wim.zwinkels@epsilon-research.nl

Table of Contents

1 Introduction

In preparation for the programming period 2014-2020 Member States are required to develop performance indicators and set targets for monitoring the implementation and performance of operational programmes (OPs). Performance indicators and corresponding targets can be set at the level of output or at the level of result. Targets shall be cumulative for 2022. Member States may in addition develop annual targets.

The Commission aims to provide methodological support to ESF Managing Authorities in setting and eventually adjusting programme performance targets for output and result indicators. This background paper summarises the main methodologies for ESF target setting and adjusting in OPs in Social Inclusion. The aim of this paper is not to provide Member States with a unique method to set targets for indicators as these methods depend on the type of indicator and the availability of data. Rather the aim is to give an overview of principles, suitable methods and pitfalls for targeting. We will avoid technical details in order to keep the number of pages restricted.

A draft of this working document has been presented to and discussed with the Member States during a learning seminar on March 14, 2013. On the basis of the discussion with Member States and the Commission it has been revised and further developed.

2 Conceptual framework for target setting

2.1 Conceptual framework

In target setting several concepts can be distinguished, although sometimes different definitions are used.[1] Figure 1 summarises the main concepts and relationships.

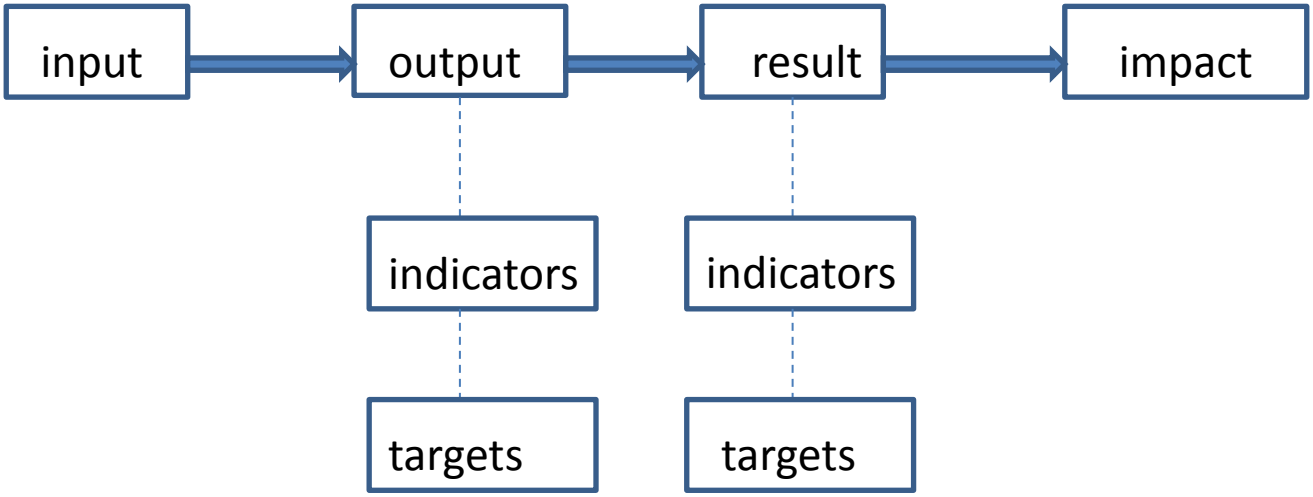

Figure 1: Main concepts and relationships in target setting

Input: The human and financial resources involved in the implementation of the programme.

Output: The goods and services, produced by the programme. Indicator and targets for outputs are number of entities or number of people undergoing the intervention financed by the programme.

Result: The effects of the intervention on directly supported individuals/entities. Results reflect the situation of the participants/entities upon leaving the intervention (immediate results) compared to the situation before, or sometime after leaving (longer-term results). Indicators are defined in terms of the goals of the programme, e.g. the employment status.
A participant/entity might not have a job directly after leaving a training programme financed by ESF. However, this may be the case after a couple of months. On the other hand, people who leave the programme having found a job might soon become unemployed again. Measuring results in longer-term indicators thus avoids focussing on short-term results that do not always proof to be sustainable.

Impact: The direct or indirect, foreseen or unforeseen, effects of the ESF on the direct participants/entities or on other groups.

[1] See Deloitte (2010), Feasibility Study of Output-Based and Result-Oriented Conditionally Systems for the European Social Fund, or EVALSED http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/guide/index_en.htm.

The various concepts can be measured by indicators. Some ESF-indicators need to be quantified by targets. The target is the value of the indicator which policy action aims to achieve. For ESF indicators and targets are restricted to outputs and results. The draft ESF and Common Provisions Regulations do not foresee the use of impact indicators in operational programme. The Commission considers that impact should be assessed in evaluations.

2.2 Aspects of target setting in social inclusion

Before describing methods of setting targets we briefly describe issues that play a role in target setting and that are relevant for choosing the right method.

Defining indicators

Targets are to be set for indicators. Sound indicators are therefore an important step in target setting. Good performance indicators are SMART:
- specific to the goals and priorities
- measurable, data should be available
- achievable in relation to current or future agreed resources
- realistic but challenging enough to be motivating
- time related to ensure that achievement of the target can be measured within a certain time frame and focus maintained.

The draft Guidance Document on Monitoring and Evaluation of European Cohesion Policy for ESF sets out these SMART criteria. For result indicators it is added that they should be linked in as direct way as possible to the operation or the person/entity supported. The High Level Group Reflecting on future cohesion policy[2] adds that indicators should be built, as far as practicable, on available underlying data, their measurement not imposing too large a burden on Member States, on enterprises, nor on citizens.

The draft ESF Regulation already contains a list of common output and result indicators. They are common because all OPs supported by the ESF are required to record and store the data for these indicators. The rationale for common indicators is to enable the collection and aggregation of data sets from Member States in order to report achievements at EU level. The number of these common indicators is kept to a strict minimum of the most important information needed to report on the main scope of ESF support. Common indicators thus represent the minimum set of indicators for each OP. They may be complemented with programme-specific indicators which can be financial, output or result indicators.

The common indicators are listed in Annex A of this paper. The list of common indicators contains three to four social inclusion output indicators: (1) migrants, people with a foreign background, minorities (including marginalised communities such as the Roma), (2) disabled and (3) other disadvantaged and (4) low skilled (ISCED 1 and 2). Other programme-specific social inclusion output indicators could for instance be: homeless participants or suffering from housing exclusion, ex-offenders, drug addicts and substance abusers, lone parents, participants from a workless household, pupils with special educational needs etc.

[2] F. Barca and Ph. McCann (2011), Outcome Indicators and Targets – Towards a Performance Oriented EU Cohesion Policy.

Each of these target groups can be linked to result indicators such as number of participants gaining a qualification upon leaving, achieving (self-)employment upon leaving or 6 months after leaving the programme, participants with an improved labour market situation upon leaving.[3]
A final issue for defining indicators to be mentioned is that they should focus on results as far as possible and avoid perverse incentives. Suppose for example that sanctions are encouraging work resumption of unemployed persons. Then it is not advisable to define indicators in terms of 'number of sanctions given' as caseworkers will then be stimulated to give as sanctions. Although it can be useful to register and evaluate the number of sanctions given, instead indicators should be defined in terms of employment status in order to avoid this strategic behaviour known as 'gaming'.

Type of target

Targets can be expressed in absolute numbers or in percentages, however percentages are only allowed for result targets. Percentages can be intuitively more attractive. However, with percentages one should keep in mind that numerator and denominator are expressed for the same target group. E.g. one has to compare the number of female participants from ethnic origin leaving a training programme with a qualification to the total number of female participants from ethnic origin in the programme and not to the total number of (female) participants in the project.

In social inclusion issues targets can also reflect reducing inequality. Poverty and income distribution can be measured by *Lorenz curves* and *Gini coefficients* (as World Bank does). Usually these measures are used in monitoring rather than target setting, however they can be an interesting source for Member States in target setting in the area of social inclusion.

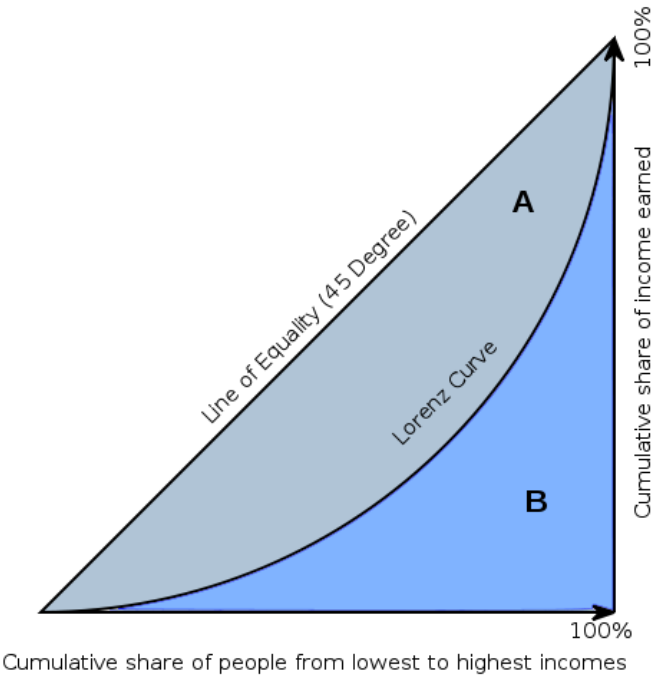

Figure 2: Lorenz curve for income distribution

[3] European Commission, DG EMPL (November 2012), Possible ESF Programme Specific Indicators, Draft Technical Document.

The Lorenz curve (Figure 2) is a graphical representation of the cumulative distribution, where it shows for the bottom x% of individuals (or households), what percentage y% of the total income they have. A perfectly equal income distribution would be one in which every person has the same income. In this case, the bottom N% of society would always have N% of the income. This can be depicted by the straight 45-degree line. By contrast, a perfectly unequal distribution would be one in which one person has all the income and everyone else has none. In that case, the curve would be at y=0 for all x<100% and y=100% when x=100%. The *Gini coefficient* is the area between the line of perfect equality and the observed *Lorenz curve*, as a percentage of the area between the line of perfect equality and the line of perfect inequality. The higher the coefficient, the more unequal the distribution is. In Figure 2 this is given by the ratio A/(A+B).

Variations to the Lorenz curve exist, e.g. in the form of quintile ratios: total income received by the top 20% of the country's population with the highest income (top quintile) to that received by the 20% of the country's population with the lowest income (bottom quintile).[4]

Challenging but realistic

Targets should be challenging enough to prevent a decline in motivation after realising the target and should not be set too high to enable the target to be reached. A necessary condition for motivational issues to play a role is that actors have access to the value of the target indicator. This is probably not always the case. However, given the target planning period in 2022 this might play a role if it becomes clear that the 2022-target can easily or almost not be reached.

Baseline setting

The draft ESF Regulation requires setting a baseline for result indicators. No baseline is required for output indicators. The ESF monitoring and evaluation guidance further specifies that only those result indicators which are linked with a target need a baseline. Baselines establish a reference value against which progress is subsequently measured. That means baselines capture a situation at the start of the programming period for the dimension covered by the respective result target. Baselines thus serve as a reference for setting targets and assessing progress. Baselines need to be expressed in the same statistical unit as the target. Baselines can be established on the basis of an existing similar programme or intervention, be it an ESF or national/regional programme. The definition of baselines differs from academic use where it usually means absence of intervention.

Adjusting targets

Member States can adjust targets in duly justified circumstances such as changes in financial allocation (output targets) or an economic downturn as we have been experiencing since some time (this might be more relevant for result targets than output targets). For instance, the collective mass lay off in a car factory may – depending on the national and regional labour market situation - have an impact on the labour situation in a region or Member State. When adjusting the targets, the managing authority should use the same methodology used when setting the targets in the first place. This calls for a robust methodology to set the targets that can be used for adjusting them as well.

[4] T. Atkinson, E. Marlier and B. Nolan (2004), Indicators and Targets for Social Inclusion in the European Union, Journal of Common Market Studies, 42 (1).

(Des-)aggregation

Targets have to be defined at the same level as the indicators, i.e. the investment priority.

One can define three approaches in defining targets:

1. Horizontal: based on values of the indicator in the past or similar programmes (baselines)
2. Bottom up: targets are defined at a lower level of aggregation and then summed-up
3. Top down: targets at a higher level are disaggregated

Targets are not only an obligation to be fulfilled but can be used as a tool for dividing available budgets to a lower level, e.g. (sub-)regions. Suppose that (sub-) regions vary in costs or in effect levels that can be reached, then this information can be used to divide or reallocate the available budget to areas with the lowest (unit) costs or the best results as we shall see.

Data

As mentioned data availability is a condition for setting targets. The availability of data is a serious challenge when it comes to defining output and result targets for social inclusion.

First of all, definitions for the disadvantaged participant groups may differ between Member States. Therefore national definitions shall supplement the definitions indicated by the Commission. This approach was chosen because either no EU-wide definition is available or imposing a harmonised definition would cause considerable administrative burden. It is recognized that this fact and also national differences in data collection will affect the data consistency across Member States in some way. However, the process of target setting is not fundamentally different, once national definitions have been set according to availability of these data. Moreover, data on common indicators and/or participant groups from several Member States can be used for future target setting. They can serve as reference values in baseline setting or incorporated in econometric modelling.

A more serious problem is data availability. Characteristics of the target group such as whether they are disabled, migrant, ex-offenders or drug addicts might be available as administrative data if these characteristics are a necessary condition to participate in the project. Otherwise surveys are needed. This might heavily affect data quality, e.g. in the case when participants have to report whether they are ex-offenders, drug addicts etc. Linking microdata of the project (or other type of activity) to register data might not always be a possibility due to privacy reasons or restrictions in data registration. When surveys are used to collect data minority groups might be difficult to reach or face language problems. Moreover, some of these characteristics are rarely available in administrative data. Hence sampling will be very difficult for certain groups such as the homeless. The main solution to this problem is to make use of data generated in previous projects where this information is registered and these characteristics are admissibility criteria.

For longer term result indicators, e.g. persons with an improved labour market situation 6 months after leaving, a limiting factor will be that project data are not always following participants after leaving the programme. In this case surveys have to take place. A point of attention is the sampling of these surveys as it can be the case that the chances of tracing persons is correlated with the indicator itself. In

particular this is the case for the homeless but also other socially isolated groups may be less willing to trace or to participate in the survey.

Another European data source is EU-Statistics on Income and Living Conditions (EU-SILC). EU-SILC is the EU reference source for comparative statistics on income distribution and social inclusion at the European level. It provides two types of annual data for 27 European Union countries, Croatia, Iceland, Norway, Switzerland and Turkey:

- Cross-sectional data pertaining to a given time or a certain time period with variables on income, poverty, social exclusion and other living conditions, and
- Longitudinal data pertaining to individual-level changes over time, observed periodically over a four year period.

SILC data are not sufficient to set targets as there is no link with programmes. There is no information on participating in active labour market programmes. Only data on following training (however not necessarily ESF funded) is available. However, SILC data can be used for reference values, to set baselines or to adjust targets.

Data in the field of social inclusion from EU-SILC that can be used:

- Income (personal and household level)
- Risk of poverty (household income below 60% of national median income, the so-called Laeken definition)
- Material deprivation (debts, expenses, surroundings)
- Living in a workless household
- No participation ALMP, only training (not necessary ESF funded)

Data for personal characteristics that are relevant in the field of social inclusion are:

- Age, gender, educational level, disability
- Nationality, country of birth, however not ethnic origin

See Irish Department of Social Protection (2011)[5] for an example of target setting in poverty making use of SILC. The information of EU-SILC is less suitable for the target setting process of other specific target groups as characteristics for the homeless, ex-offenders and ethnic minorities are not directly available. However, some approximations are possible. EU-SILC contains an ad-hoc module (only for the year 2008) covering the labour market situation of migrants which can be very useful for target setting on migrants.

For adjusting targets also studies can be used, e.g. Ecorys (2011)[6] where OECD and Eurostat data are used to estimate the influence of the business cycle.

[5] Irish Department of Social Protection (2011), Synthesis Report on the EU Peer Review on the Setting of National Poverty Targets.

[6] Ecorys (2011), Performance Targets for ESF Operational Programmes.

3 General methods for setting and adjusting targets

In this section we describe methods that can be used for setting and adjusting output as well as result targets. We will distinguish qualitative and quantitative methods. The survey on target methodologies shows that most Member States use more than one method. In almost all target settings qualitative methods play a role. In particular qualitative methods are used to set baselines based on experience and past performance. Qualitative methods can also be used at the end of the process to 'verify' results from quantitative methods ('reality check'). Qualitative methods are usually based on some quantitative information, e.g. per capita costs of an intervention. Qualitative methods are more likely to be support than to replace each other and a combination of them or with a quantitative method is the most likely approach. For quantitative methods it is more likely that one chooses the most appropriate one.

We will describe each method, mention advantages and disadvantages, show which data are needed and give some examples. In theory every method for setting targets can also be used to adjust targets, however the reverse is not necessarily true.

3.1 Qualitative methods

Expert opinions

One of the most often used methods is expert opinions that are used to estimate a target or a baseline. In the absence of sound data for estimating this method might even be the only way to set a baseline or target. Expert opinions can be determined in focus groups or through questionnaires. A disadvantage of focus groups might be the tendency of a dominant group member to exert influence the opinion of the other group members. Expert opinions usually introduce a 'normative' element, in particular at the end of the target setting or target adjusting process.

Delphi method

A specific qualitative method is the Delphi method. In the Delphi method experts give their estimates in two or more rounds (mostly this is a focus group approach). After each round, a facilitator provides an anonymous summary (or average) of the experts' forecasts from the previous round, which can be supplemented by reasons they provided for their judgments. In the next round experts are encouraged to revise their earlier answers in light of the replies of other members of their panel. It is believed that during this process the range of the answers will decrease and the group will converge towards the "correct" answer. Finally, the process is stopped after a pre-defined stop criterion (e.g. number of rounds, achievement of consensus, stability of results) and the mean or median scores of the final rounds determine the results.

Delphi is based on the principle that forecasts from a structured group of individuals are more accurate than those from unstructured groups ("collective intelligence").

A systematic review of 25 empirical studies comparing Delphi study with standard interacting groups concluded, with some caution, that Delphi groups outperform other types of groups in decision making and forecasting. [7]

In general:
The advantages of expert opinions and the Delphi method are:

- no data are required
- quick and the costs are relatively low
- the method itself can create some support for acceptance to experts and public as experts are given the opportunity to estimate the target
- it supports and checks quantitative data at the end of the process

The main disadvantages of expert opinions and the Delphi method are:

- less transparency in the way targets are motivated
- targets can be in an undesirable way influenced by experts who are dominant or have a hidden agenda
- less exact and less reliable than quantitative data

Use of reference values

Literature study or simple data counting can lead to reference values that can be used as a baseline or a target. These can be historical data of the programme or data from similar programmes or projects in other regions or Member States. Please note that if calculation or modelling takes places with these data we will consider them to be benchmarking methods or other quantitative methods.

The main advantage of using reference values is that they:

- are easily accessible
- are transparent
- can be sound from (high quality) studies

Though data might be available, a problem in using data from other programmes or historical data from the same programme can be the comparability to the operational programme. Disturbing factors can be:

- other characteristics of the target group
- different economic situation
- differences in legal and institutional context

3.2 Quantitative methods

Trend analysis

If we only have figures for the indicator values we can apply trend analysis to calculate a target value, which can be used as a baseline value or a target. With trend analysis we try to estimate a trend term in values for indicators. These can be applied to the Operational Programme as a whole or also for priorities.

The simplest curve is the straight line. For example let C(t) be the unit costs for a training per participant or the result indicator of participants with employment upon leaving in year t then:

[7] G. Rowe and G. Wright (1999), The Delphi Technique as a Forecasting Tool: Issues and Analysis, International Journal of Forecasting 15.

C(t) = a + b*t

where:
a = autonomous part (constant)
b = trend factor

The trends can be relatively easy used for forecasting.
In case of the linear trend:
C(t+1) = a + b*(t+1) = C(t) + b

If there are seasonal data one can apply seasonal adjustments. Furthermore, it is advised to check for outliers as they can influence results to some extent. Financial data can be corrected for inflation in order to get a smooth series.

Example:
Suppose that we have the following series for unit cost, adjusted for inflation:

	2007	2008	2009	2010	2011	2012	2013
unit cost	100	98,4	97,3	96,1	95,5	94,8	94

In Excel we can estimate several shapes. A linear trend will fit well, and there are no outliers. Every year costs will be 0,96 Euro less as the estimate shows. The line fits well over the whole period, so it is to be expected to continue. The estimated unit costs for 2014, corrected for inflation, can be calculated as 94-0,96=93,04 etc for 2015 and so on.

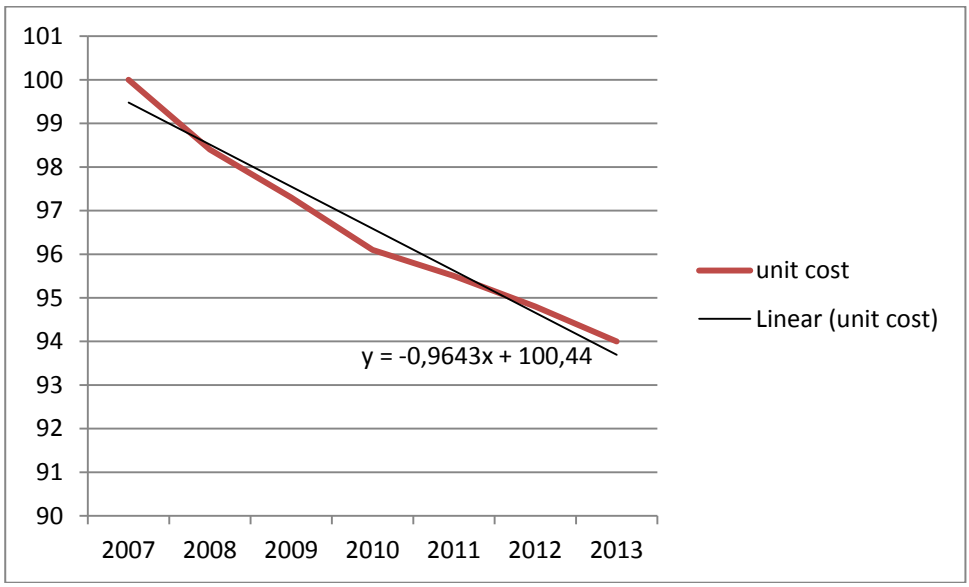

Figure 3: Example of estimating a trend line

Other possible specifications are:

The exponential curve:
C(t) = exp(a + b*t) so that log C(t) = a + b*t

The parabolic curve:
C(t) = a + b*t + c*t^2

The logistic curve has:
C(t) = 1 / (a + b*r^t) with 0< r < 1
This logistic curve has the property that future estimates will lay between 0 and 1, which is a condition if we are setting targets for indicators defined as 'shares' (percentages divided by 100).

Further extensions are possible in the form of (a combination of) autoregressive models and moving average models.[8]

Whatever extrapolation method is chosen, one has to take care not to groundlessly use them for targets. In particular with linear trends the risk is that forecasting will create unrealistic outcomes or even become 'out of bounds', e.g. predictions of > 100% if the forecasting period extends. Setting targets in this way is only valid if the drivers for change will be sufficient to maintain the performance increase that has already happened and provide an additional effect.

Advantages:
- relatively easy
- data on indicator (and probably some underlying data) suffice

Disadvantage:
- Rather mechanic, changes in target values are not explained

Economic optimum

In the economic optimum method the target is calculated as the weighted average of underlying subsets. These can be regions, type of activities or subgroups according to characteristics of the target group. The underlying idea is not only to calculate targets for these subgroups but to invest (divide budgets) making use of the differences in expected outputs/results, e.g. by minimizing the share of regions with higher unit costs/lower results and maximising regions with lower unit costs/higher results. Boundaries for these solutions can be set, to reflect minimum spending or minimum number of participants. The problem then turns into a (in most cases) linear programming model and Excel or any suitable software package (see section 6) will find a solution which can be used as target.

An example:
Suppose the results of a specific objective 'promote the establishment of start-ups by ethnic minorities' with a result target 'sustainable start-ups after 6 months' over period T0 (e.g. ESF 2007-2013) are given in table 4 for 4 regions. For the new period T1 (e.g. ESF 2014-2020) budgets may differ up to 10 percentage points. Assume that the result scores of the different regions remain the same. In this simple example 10% of the budget will be re-allocated from regions with the lowest score (1 and 3) to regions 2 and 4. The target (or the baseline if this method is used for baseline setting) for the new period will become 57% compared to 53% over period T0.

[8] For further reading see e.g. C.W.J. Granger (1980), Forecasting in Business and Economics.

	Region 1	Region 2	Region 3	Region 4	Total
T0					
% start-ups	**40**	**70**	**50**	**60**	
share of budget	**0.3**	**0.1**	**0.2**	**0.4**	
target (total)					**53**
T1					
% start-ups	**20**	**40**	**20**	**50**	
share of budget	**0.2**	**0.2**	**0.1**	**0.5**	
target (total)					**57**

Table 4: Example of economic optimum method

Shift-share analysis

Related to the economic optimum method is the shift-share analysis. In a shift-share analysis the changes in a series of an output or result indicator are decomposed in changes in the relative shares of subgroups and changes in the indicator value for different subgroups.

Suppose I is the value of the overall indicator, I_i and w_i the relative shares we can write:

$$\Delta I = \sum_i (w_i \Delta I_i + \Delta w_i I_i + \Delta w_i \Delta I_i)$$

The shift-share analysis is used by OECD and World Bank in particular for poverty and income distribution indicators. OECD[9] decomposes changes in poverty rates into three components:

- the part due to changes in market-income poverty for each of several groups within the two household types, while keeping constant both the structure of the population and the effect of taxes and transfers in reducing poverty for each group;
- the part due to changes in the effect of taxes and transfers in reducing market-income poverty for each group, for a given population structure and market-rate poverty for each group; and
- the part due to changes in the structure of the population by both household type and number of workers in each household, for a given market-income poverty rate and level of effectiveness of tax and transfers in reducing poverty in that group.

These results can be used in setting a baseline or target by using the historical values applied to new weights or by applying historical weights to new calculated targets by applying other methods (e.g. trend analysis). The use of this method is also suitable for target adjustment when circumstances change.

For example, target values for a programme for migrants with result indicator ´people in employment upon leaving the programme´ have been set in the past in 4 regions, see table 5. Due to the economic crisis result targets have to be adjusted for T1. Furthermore, developments in the inflow of migrants differ between regions compared to the situation the target was set T0.

[9] OECD (2008), Growing Unequal, Income Distribution and Poverty in OECD Countries.

	Region 1	Region 2	Region 3	Region 4	Total
T0					
% empl	**40**	**50**	**50**	**60**	
share	**0.3**	**0.1**	**0.2**	**0.4**	
					51
T1					
% empl	**20**	**40**	**20**	**50**	
share	**0.1**	**0.2**	**0.1**	**0.6**	
	2	**8**	**2**	**30**	**42**

Table 5: Example of shift-share analysis

In this situation the total change ΔI is -9, which can be decomposed in:

- change due to changing result indicators $\sum_i w_i \Delta I_i$: -17
- change due to changing shares of regions $\sum_i \Delta w_i I_i$: +4
- Cross product $\sum_i \Delta w_i \Delta I_i$: +4

Note that if the result indicators for the regions remain unaffected (ΔI_i =0 for all i) the formula is reduced to the economic optimum method.

Advantages:

- analysis shows the underlying analytical path to go from baseline to target or target to new target
- no microdata are necessary (although 'cells' might be based on microdata)
- simple to use and understand

Disadvantages:

- underlying target setting/adjusting necessary
- no causality included

4 Specific methods for setting output targets

Unit cost

A commonly applied method for setting output targets is the unit cost method. Historical unit costs for a certain output indicator, e.g. number of participants, are calculated by dividing the number of participants by the allocated budget in the past. This results in the costs per unit (participant). Given the total of the allocated budget of the programme in the period 2014-2020, one can calculate the number of participants by dividing the allocated budget by the historical unit cost.

The same disturbances mentioned above can be of influence when setting a target. Additionally there may be the following factors to include in setting a target:

- price inflation should be accounted for
- fixed costs can be higher or lower, affecting the unit cost price, in particular when continuing an existing project
- (increasing or diminishing) returns can play a role, i.e. the 'output' function depends on the scale or on learning effects
- prices differ between regions, sectors, characteristics of the target group, a change in the distributional 'mix' of these characteristics in the new Programme can change the unit costs for the target.

Advantages:

- simple

Disadvantages:

- no correction for exogenous factors

Benchmarking

Benchmarking methods usually use reference values which are corrected for differences in the situations that are compared. For output targets these differences can be characteristics of the target population and organisational factors that influence the relationship between inputs and outputs. In particular in social inclusion costs of supporting people may differ between target groups, e.g. think of severity of the disability. For result targets benchmarking methods can also be used, however it is less common if econometric solutions are possible.

A particular benchmarking method is Data Envelopment Analysis (DEA). DEA is a linear programming based technique for measuring the relative efficiency of a fairly homogeneous set of 'decision making units' (can be regions in ESF-target setting) in their use of (multiple) inputs to produce outputs. It identifies a subset of efficient 'best practice' units and for the remaining units the magnitude of their inefficiency is derived by comparison to a frontier constructed from the 'best practices'. DEA derives a single summary measure of efficiency for each unit based on the ratio between outputs and inputs.

$$Efficiency = \frac{Weighted\ sum\ of\ outputs}{Weighted\ sum\ of\ inputs}$$

For the inefficient DMUs, DEA derives efficient input and output targets and a reference set (or peer group), corresponding to the subset of efficient DMUs to which they were directly compared.

For a simple example, see Figure 5. Assume 4 regions (A, B, C, and D) that use a certain amount of the budget (input) and reach some output level e.g. number of participants. The envelopment surface will differ depending on the scale assumptions that underpin the model. Two scale assumptions are generally employed: constant returns to scale (CRS), and variable returns to scale (VRS). CRS reflects the fact that output will change by the same proportion as inputs are changed (e.g. a doubling of all inputs will double output); VRS reflects the fact that production technology may exhibit increasing, constant and decreasing returns to scale. The frontier defines the full capacity output given the level of fixed inputs. With constant returns to scale, the frontier is defined by region C for all points along the frontier, with all other points falling below the frontier (hence indicating capacity underutilization). With variable returns to scale, the frontier is defined by regions A, C and D, and only region B lies below the frontier i.e. exhibits capacity underutilization. For CRS the underutilization for region B=O3-O1 for VRS O2-O1.

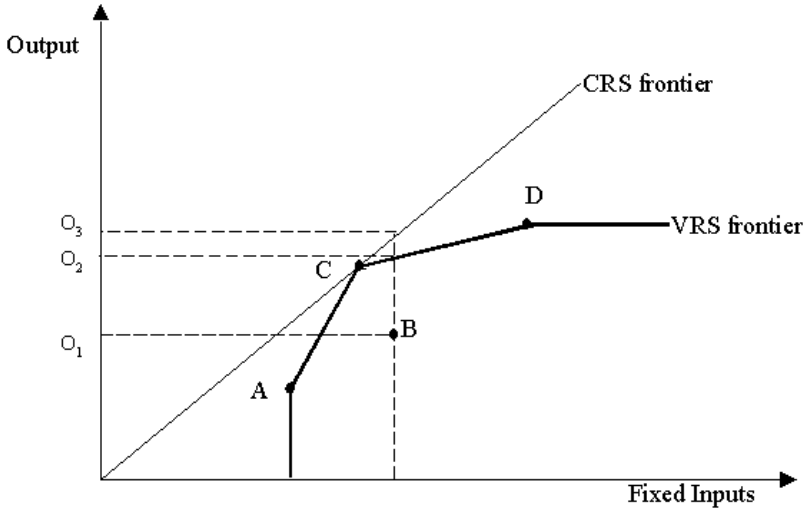

Figure 5: Example of Data Envelopment Analysis method

A related method is stochastic frontier analysis (SFA). In fact this is the DEA with parameterization. With y representing output and x input, the basic specification for SFA is:

$\log y(i) = \beta 0 + \beta 1 \log x(i) + v(i) - u(i)$

V(i) represents an error term and u(i) the underutilization for region i.[10]

Advantages:

- Gives insight in differences between units
- Graphical illustration possible (if not too many in-/outputs)

Disadvantages:

- Rather technical (in particular SFA)

[10] For further reading see the main article D.J. Aigner, C.A.K. Lovell, P. Schmidt (1977), Formulation and Estimation of Stochastic Frontier Production Functions; Journal of Econometrics, 6:21–37.

5 Specific methods for setting result targets

5.1 Issues in target results for social inclusion

The main objective of result indicators is to capture the (expected) effects on participants or entities brought about by the operational programme. Though for some of the output indicators relatively simple methods as the unit cost method give reasonable results this is usually not the case for result indicators. In general two disturbing factors play a much bigger role: selection effects and non-programme factors.

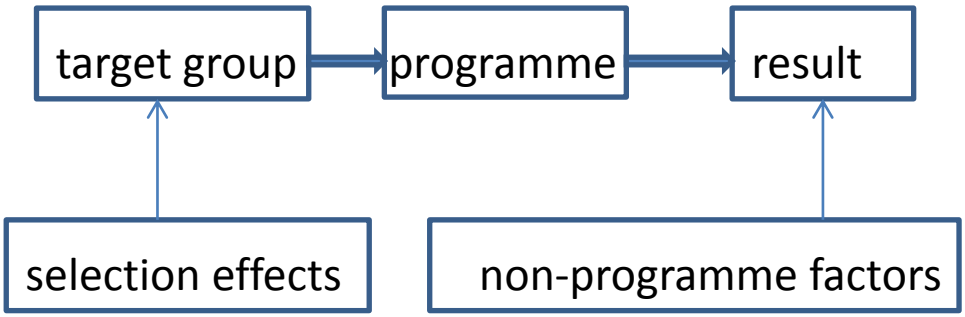

Figure 6: Disturbing factors in setting result targets

Selection effects
Results are usually influenced by characteristics of participants. For example, it will be more difficult to integrate people with a severe disability into the labour market than people who only have a minor impairment. Selection effects (by local authorities or by participants themselves by participating or not) can change the distribution of characteristics. One has to avoid that defining the indicators and targets leads to 'creaming' or 'cherry picking', i.e. selecting the most favourable participants to achieve the results. These can even be individuals who don't benefit at all from the programme ('deadweight loss'), e.g. unemployed person who would have found a job without the programme.

Some of these characteristics can to a certain extent be 'internalized' by changing the definition of the indicator. One can think of separate targets for persons depending on the severity of the handicap. Please note that this is not possible for common indicators. However, some common indicators have already been broken down for different target groups e.g. disabled persons, ethnic minorities and other disadvantaged groups.

It can also be the case that selection is to be made by external factors. Budgets have to be divided amongst certain regions with each a different distribution of ethnic minorities (first, second generation) with different chances on the labour market or different patterns in school leaving. Or equal opportunities play a role, which means that every ethnic minority group has to receive a certain share of the budget. In these cases the solution is to adapt result target levels according to the characteristics of the group (if we have this type of information).

On the contrary, selection is in some circumstances desired. We want to spend our budget on participants who benefit from the actions undertaken. For example selection on the basis of motivation of the client can considerably increase results. Screening and diagnosis instruments can be used prior to the programme. In this case one would like to reflect the desirable selection effects in the indicator and target.

So basically there are three approaches to selection effects:

1. internalize them in the definition of the indicator to avoid creaming
2. adjust result target levels, accounting for heterogeneity of the groups
3. do nothing if selection is favourable to avoid interventions to people who will likely not benefit from the intervention

Non-programme factors

Although the advice is to define indicators and set targets as close to the performed activities as possible, the result indicator can still be affected by non-programme factors. In particular this is the case for employment type of indicators that are affected by the business cycle. Also differences in the legal and institutional context can lead to different results.

Ideally, for a good baseline for a result indicator we would like to know:

- what are differences between effect sizes of (homogeneous) groups or projects/interventions
- what has been the influence of disturbing factors on the reported value for result indicator

For formulating targets we ideally need to quantify:

- the changes that occur in characteristics of the target group and in the non-programme factors
- the change in effect parameter
- calculate the resulting effect

We will see this in the next sections when discussing the various methods.

Net effects

Please note that selection effects and effects of non-programme factors are important to get a view on *net* effects of investment, though measuring impact remains the area of monitoring and evaluation rather than target setting. Economic optimization of results (i.e. net impact of ESF-investment) is usually but not always equal to maximization. Take a simple example where with intervention in region 1 80% finds a job and without intervention 60%. Suppose that in region 2 these figures are 70% with intervention and 40% without intervention then maximising the effects will favour region 1 whereas maximising the *net* effects clearly points at region 2 as the most effective.

5.2 Macro-econometric approach

Time series modelling

In a time series model we add causal factors that have a relationship with the indicator we want to set a target for. The advantage of this is that we can attribute part of the variation in the past to external factors and use this information for

forecasts. Moreover, using this method gives additional proof to the new target and decomposes differences in targets.

The most important external factor to correct for in social exclusion is the business cycle affecting the (local) labour market situation. It is widely known that disadvantaged groups on the labour market such as disabled, ethnic minorities and people with low education suffer in a disproportionate way from an economic downturn compared to 'normal' employees. As a consequence the returns from investment by ESF-funded activities are diminished if the business cycle slows down and result target values should be set lower.

An obstacle for estimating time series models in social inclusion is the relatively short time period. In order to estimate the relationship, one would preferable having monthly or quarterly data or data of various regions, Member States to have enough observations.

Another challenge is to find consistent and reliable time series data for indicators referring to social inclusion performances for a variety of countries and time periods. This impedes the creation of good panel data sets.[11]

Advantages:
- changes in target values are explained
- the same method can be applied for adjusting targets in a transparent way

Disadvantage:
- can be rather complex
- possibly not enough (time) observations

Panel data model

Comparable time series on interventions can be taken together. These can be time series on the same type of programme for different regions within a Member State or for several Member States. This type of data is called panel data. This approach differs from the ideal situation in the sense that the effectiveness itself is not measured. At most, differences in effect sizes between Member States or regions are measured. Based on macro-economic programme data we can estimate target levels. We can make use of special estimation techniques such as *fixed effects* or *random effects* and also allow for time specific correlations and heteroscedasticity. For these topics we refer to econometric textbooks on panel data.[12]

Advantages:
- changes in target values explained
- data are available for this approach

Disadvantages:
- Member State specific effects or different definitions make figures incomparable

Adjusting targets

Macroeconomic model can also be used to adjust targets. This approach has been followed by Ecorys, though not in the field of social inclusion.[13]

[11] Ecorys (2011), Performance Targets for ESF Operational Programmes, p.89.

[12] E.g. A. Cameron and P. Trivedi (2009), Microeconometrics Using Stata.

[13] Ecorys (2011), Performance Targets for ESF Operational Programmes.

The employment level of Member States (EMPL) is modelled as a linear function of specific Member State characteristics (X) such as expenditures on active labour market policies and gross domestic product (GDP) as the business cycle indicator:

EMPL (MS,t) = b0 + b1 X + b2 GDP (MS,t)

The model is estimated on Eurostat and OECD data. It is concluded that one percentage point higher GDP roughly gives one percentage point higher employment level in the long run. However, this is a general number, and does not say much about disadvantaged target groups.

5.3 Micro-econometric approach

The micro-econometric approach tries to capture all relevant effects that play a role in targeting. This starts with a detailed analysis to determine all effects over the past period including the effects of the programme. Ideally we would like to know what would have been the result without the ESF support. However, if a person enters a programme the result without that programme will be unknown by definition (in literature this is known as the counterfactual). There are several ways to determine the effectiveness of the programme depending on the data and practical aspects of the programme.[14]

Social experiment
Scientifically, the most preferred setup is a social experiment with a randomised control group. Random assignment is important as there will be by definition no selection effects or differences in population characteristics. The result indicators of the experimental and control group can be compared directly. Differences show the effect of the experiment. In practice social experiments are hardly carried out as it will often be costly to create a 'controlled' environment. Sometimes it is also argued that it is unfair and unacceptable to exclude possible participants in the control group from policies. So usually this is not a source of information at the project and it will even be rarer at programme level.

Discontinuity approach
The discontinuity approach makes use of the fact that some interventions use certain criteria for target groups. Information is then gathered for a control group of people who are just not eligible to receive the intervention. For example, suppose that a certain project is only available for unemployed persons from 50 years and older. It is then possible to create a control group of unemployed persons of exactly or just under 49 years. However, this still requires that there are no selection effects for the target population as in that case correction is needed and a discontinuity approach doesn't have advantages. The practical usability of this method is therefore very restricted.

Matching
Matching consists of composing a control group consisting of individuals who have (roughly) the same characteristics as the intervention group. For every person that has entered a project a 'buddy' (or more) has to be found in the data. An important

[14] For a more extended overview see: European Commission, Design and Commissioning of Counterfactual Impact Evaluations, 2012.

condition is that all characteristics that affect the result indicator need to be included as matching characteristics. Usually for interventions in the social inclusion field there are a lot of characteristics and this is a relative time-consuming process. Moreover, for every intervention this needs to be done or available and to be aggregated to the programme level.

Quasi-experimental methods
The most common used method is a quasi-experimental design. A micro econometric model is developed with the result indicator at the individual level as the dependent variable.
In the model are (or can be depending on availability of data) included:

- personal characteristics
- non-programme factors, including the business cycle, labour market and other social factors
- (type of) intervention

The specification of the model depends on the type of indicator:

- logistic regression model or linear probability model for (0,1) variables, e.g. in employment after the intervention yes/no
- duration model if indicator needs to reflect time till job is found

As all data are available, targets can be relatively easy computed:

- estimate the change in personal characteristics of the target group
- make use of (external) forecasts of business cycle, labour market developments and other
- make assumptions on the effect parameter (remain same or increase effectiveness)
- fill-in the new values in the existing model and calculate on the micro level the consequences for the result indicator
- aggregate this to programme level

This approach has been followed in Bartik et al. (2009)[15] and Ecorys (2011)[16]. However both applications are in the field of employment and are applied to target adjusting rather than target setting.

Bartik et al. use quarterly data for several indicators (employment, earnings, attainment of degree, literacy) at area level.

Y (i,area,t) = B0 + B1 * X (i,area,t) + B2 * D (area,t)
Y = indicator value
X = individual characteristics (age, sex, ethnic origin, educational level, disability, etc)
D=local unemployment rate
B0, B1, B2: estimated parameters

B2 is the estimated elasticity which reflects the change in indicator as a result of a change in local unemployment area.

[15] T. Bartik, R. Eberts and W. Huang (2009), Methodology for Adjusting GPRA Workforce Development Program Performance Targets for the Effects of Business Cycles.
[16] Ecorys (2011), Performance Targets for ESF Operational Programmes.

New targets are rather simply calculated from the model outcomes by using the old target and adding (or subtracting if negative result) the elasticity multiplied by the change in local employment rate in the area.

Advantages:

- all relevant effects are quantified, a baseline can be relatively easy computed and the results can also be used to set targets

Disadvantages:

- requires individual data and characteristics over time, which might not be (easily) available
- knowledge of using the right econometric techniques and use of statistical packages is necessary
- the final target remains more or less a 'black box' for non-specialists

Ecorys (2011) has developed a similar model applied to the Flemish labour market. Elasticities are a bit smaller than in Bartik et al. Though applied to employment, Ecorys has also estimated coefficients for certain subgroups that can be used in the context of social inclusion policies. It is concluded that older people (50+) and low-educated react the same to changes in the labour market situation (regional unemployment rate). Non-nationals however are less affected by the regional labour market in finding a job. Ecorys further suggest using EU-SILC data to estimate relationships and elasticities in the area of social inclusion.

Ecorys believes that the results are robust enough to be used as an indication for the level of adjustment of ESF-targets. For more precise target adjusting and target setting a similar analysis should be carried out in order to correct for different mechanisms and specific labour market factors as functioning of the labour market, institutional and cultural factors.

6 Statistical software

In general for methods that are not too complex Excel will usually be sufficient, provided the data are on the aggregate level (Member State, different Member States, regions, cohort).

Trend analysis and shift-share analysis can rather easily be done. For linear programming to find economic optimum Excel will also do the job, though if there are many equations for restrictions related to minima or maxima for spending budget shares to e.g. regions this might be a little time consuming.
The Regression Tool can be used to estimate the correlation of one or more dependent variables to a dependent variable. It can provide the intercept and slope coefficients to "draw the line" for current and future data points. Simple time series are therefore possible to estimate with Excel.

If methods are more complex Excel will be insufficient. This is the case for more advanced time series models including autocorrelation, models for panel data. For microdata models always other software packages are necessary. The most common ones are SAS, STATA and EViews.

7 Concluding remarks

Target setting in the area of social inclusion is quite challenging. Hardly any study can be found on methods that have been applied to the field of social inclusion and data are not easily available.

In general, the situation for output indicators is usually less complicated. The unit cost method provides usually good results. For result indicators, the least difficulties may arise for indicators that refer to employment situation of disadvantaged groups such as lower educated, disable people. Data availability restricts the possibilities to set adequate targets for 'other' disadvantaged groups and areas (poverty, homelessness).

As suitable methods depend on the availability of data, which can also be quite challenging for social inclusion, we will conclude with an overview. For output and result indicators the most suitable method is proposed, depending on the availability of data.

Output indicators

A. Unit cost from historical data available -> yes
- Correction for inflation (base year)
- Differences in characteristics population? -> correction: shift-share -> needed number of participants per characteristic and costs
- Additional: check for trend (other than inflation)
- Benchmarking methods to raise target levels at lower level, e.g.

B. Unit cost from historical data -> no
- Other reference values indicator -> yes
    - Correction if possible for differences
- Other reference value indicator -> no
    - Qualitative methods

Result indicators

A. Historical data available-> yes
- Creaming effects -> choose from
    - internalize in definition of indicator
    - adjust result target levels, accounting for heterogeneous groups
    - do nothing if selection is favourable
- Microdata available -> micro-econometric approach
- Microdata not available
    - Data available from other areas / Member States -> yes
        - macroeconomic panel data approach
    - Data on labour market /business cycle ->
        - timeseries modeling
    - Data available from other areas / Member States -> no
        - trend analysis

B. Historical data available-> no
- Other reference values indicator -> yes
    - Correction if possible for differences
- Other reference value indicator -> no
    - Qualitative methods

Adjusting targets

A. Macro- or microeconometric model with business cycle -> yes
- apply coefficients from model

B. Macro- or microeconometric model with business cycle -> no
- Otherwise use elasticities from other studies (if transferable)

A List of common indicators

In this annex a list of common indicators is presented. The list of common indicators is copied from the annex to the draft ESF regulation:

Further definitions can be found in the draft Guidance Document on Monitoring and Evaluation of European Cohesion Policy for ESF.

(1) Common output indicators on participants
Participants refer to persons benefiting directly from an ESF investment and who can be identified and asked for their characteristics, and for whom specific expenditure is earmarked. Other beneficiaries should not be counted as participants.
• unemployed, including long-term unemployed
• long-term unemployed
• inactive
• inactive, not in education or training
• employed, including self-employed
• below 25 years
• above 54 years
• with primary (ISCED 1) or lower secondary education (ISCED 2)
• with upper secondary (ISCED 3) or post-secondary education (ISCED 4)
• with tertiary education (ISCED 5 to 8)
• migrants, people with a foreign background, minorities (including marginalised communities such as the Roma)
• disabled
• other disadvantaged
The total number of participants is calculated automatically on the basis of the output indicators. These data on participants entering an ESF supported operation are to be provided in the annual implementation reports. All data are to be broken down by gender.

(2) Common output indicators for entities
• number of projects fully or partially implemented by social partners or non-governmental organisations
• number of projects targeting public administrations or public services
• number of micro, small and medium-sized enterprises supported
These data are to be provided in the annual implementation reports.

(3) Common immediate result indicators on participants
• inactive participants newly engaged in job searching upon leaving
• participants in education/training upon leaving
• participants gaining a qualification upon leaving
• participants in employment upon leaving
These data are to be provided in the annual implementation reports. All data are to be broken down by gender.

(4) Common longer-term result indicators on participants
• participants in employment 6 months after leaving
• participants in self-employment 6 months after leaving
• participants with an improved labour market situation 6 months after leaving

These data are to be provided in the annual implementation reports. They are to be collected based on a representative sample of participants within each priority axis. Internal validity of the sample should be ensured in such a way that the data can be generalised at the level of priority axis. All data are to be broken down by gender.