Risk Profiling: A Primer

Emil D Tesliuc & Vlad Grigoras
World Bank

Risk Profiling: A Primer

- **Why** inspections based on risk-profiling
- **How** to do inspections based on risk-profiling
- **Practical aspects** of doing inspections based on risk-profiling

Untargeted social inspections…

Assume that:

- Error fraud and corruption (EFC) occurs in roughly 5% of cases and closer to 10% when benefits are more complex (e.g. means-tested)
- The average benefit in Romania is USD100
- We inspect 1,000 beneficiaries, cost per inspection = USD10

Cost	USD10*1000 inspections	USD 10,000
Benefit	USD100*50 cases of fraud	USD 5,000
Cost-benefit ratio		10,000: 5,000 = 2: 1

- The benefits are lower than the costs
- 95% of beneficiaries are compliant, and will be inconvenienced by the activity of social inspectors

Solution: Move to risk-based investigations

Targeted social inspections…

Assume that:

- We use a risk profiling algorithm
- We increase the probability of detecting EFC at 50% of the total number of social inspections

Cost	USD10*1000 inspections	USD 10,000
Benefit	USD100*500 cases of fraud	USD 50,000
Cost-benefit ratio		10,000: 50,000 = 1: 5

We need to find ways to target the social inspection campaigns on the beneficiaries with a higher risk of EFC.

Potential Solutions to target social inspections

Data matching	**Events of noncompliance that have already occurred are identified.** It is problematic when: ❑ There are no other databases to match (lack of protocols) or the databases do not contain information relevant for detecting fraud ❑ Different databases measure different things or there is no unique ID
Fraud referral	**A hot line where people can uncover cases of fraud/error.** ❑A dedicated team needs to be in place. ❑Success depends on local social values/norms. ❑A long term investment in information campaigns is needed. ❑Only some types of fraud/error can be uncovered.
Manual screening	**Manual selection of cases by inspectors based on their own knowledge of beneficiaries' behavior and environment.** ❑It increases the risk of corruption. ❑Social inspectors can miss some aspects of noncompliance.

Alternative solution: Move to risk-based investigations

What is Risk Profiling?

Set of statistical procedures that would allow the social inspectors to better identify the cases with a higher probability of EFC

Risk-scoring techniques comparable to those used to categorize clients in banking or insurance.

Advantages:

- It increases the cost-effectiveness of social inspections
- It reduces the length of time fraud stays in the system
- It decreases the number of inspections for compliant beneficiaries

Outline

- **Why** inspections based on risk-profiling
- **How** to do inspections based on risk-profiling
- **Practical aspects** of doing inspections on risk-profiling

Steps when doing inspections based on risk profiles

1. Build a dataset with the population of beneficiaries and their characteristics
2. Select a random sample of beneficiaries
3. Perform inspections on the random sample of beneficiaries
4. Identify the cases in the sample with higher probability of EFC based on their characteristics.
5. Identify in the total population the cases having the characteristics that flagged a higher risk of EFC
6. Perform inspections primarily on the cases in the population that show higher risk of EFC
7. Review the model based on new iterations

1. Build a dataset with the population of beneficiaries and their characteristics

Population of beneficiaries

No	Area	No of members	etc.
1	Urban	2	
2	Rural	3	
3	Rural	1	
4	Rural	4	
5	Rural	7	
6	Urban	2	
7	Urban	4	
.....			
.....			
100,000	Rural	1	

Types of characteristics: likely to predict fraud/error and available for each beneficiary

Examples:

- Residential area/type of locality
- No of members/no of children
- No of members of active age
- Maximum level of education
- Amount of declared incomes
- Type of family (lone parents etc.)
- Period in the program
- Health status
- Information at local level

2. Select a random sample of beneficiaries

Population of beneficiaries

No	Area	No of members	etc.
1	Urban	2	
2	Rural	3	
3	Rural	1	
4	Rural	4	
5	Rural	7	
6	Urban	2	
7	Urban	4	
.....			
.....			
100,000	Rural	1	

Sample of beneficiaries

No	Area	No of members	etc.
1	Rural	1	
2	Rural	4	
3	Urban	5	
4	Rural	2	
.....			
1,000	Rural	1	

Samples to assess the level of EFC and calibrate the risk-based tools do not need to be substantial to be use

3. Do inspections on the random sample of beneficiaries

Sample of beneficiaries

No	Area	No of members	etc.	Result of the investigation
1	Rural	1		Non-fraud
2	Rural	4		Non-fraud
3	Urban	5		Fraud
4	Rural	2		Non-Fraud
.....	...	...		...
1,000	Rural	1		Fraud

+ Input the results of the social investigation in the database of sampled beneficiaries.

4. Identify the cases with most probability of EFC based on their characteristics

The simplest method: Tables

	% in total population	% of fraudsters
Urban	70	6
Rural	30	20
Total	100	10

No of members	% in total population	% of fraudsters
1	10	2
2	40	5
3	30	10
4+	20	25
Total	100	10

	No of members	% in total population	% of fraudsters
Urban	1	7	1
	2	30	3
	3	25	8
	4+	**8**	**14**
Rural	1	3	5
	2	10	10
	3	**5**	**20**
	4+	**12**	**32**
Total		100	10%

5. Identify in the total population the cases having the characteristics that flag a higher risk of fraud

Population of beneficiaries

No	Area	No of members	etc.
1	Urban	2	
2	**Rural**	**3**	
3	Rural	1	
4	**Rural**	**4**	
5	Rural	7	
6	Urban	2	
7	**Urban**	**4**	
…..			
…..			
100,000	Rural	1	

6. Do inspections primarily on the cases with higher risk of EFCin the population.
7. Review the model based on new iterations.

Profiles are not always so clear-cut…

Most of the times, the available variables do not predict clearly the beneficiaries committing fraud

We need to use more complex statistical techniques and more characteristics of the beneficiaries to predict the probability of EFC

Examples of techniques:

- Classification trees
- Logistic regressions
- Linear regressions
- Cluster analysis

Logistic regression

It identifies the key beneficiary characteristics that contribute to whether or not a case is fraudulent or in error.

Weights each characteristics according to its importance in identifying irregularity to provide an overall risk score for each case.

The risk score can take a value of between 0 and 1 (with 0 not at all likely to be in error and 1 more likely to be in error)

Example of a logistic regression model

Population of beneficiaries

No	Family	Area	No of members	Education	Probability of fraud	Risk of fraud
1	Lone parent	Rural	5	Tertiary	0.9	**High**
2	Both parents	Urban	3	No education	0.1	Low
3	Lone parent	Rural	2	No education	0.3	Low
4	Both parents	Rural	5	Tertiary	0.7	**High**
5	Both parents	Rural	2	Secondary	0.5	Medium
…..						
100,000		Rural	1			

Outline

- **Why** inspections based on risk-profiling
- **How** to do inspections based on risk-profiling
- **Practical aspects** of doing inspections on risk-profiling

Resources needed:

- Database with beneficiaries and their characteristics
- Social inspectors using the results of the risk-profiles and provide input in the feedback loop
- A system to input the results of the social inspections in the database of the beneficiaries' sample
- Statistical team of 3-5 persons experienced with data management, sampling techniques, and inferential statistics

- **Risk-based inspections can be implemented successfully even in environments with limited technical resources.**