

# Building a Protection Vulnerability Formula



JRC Community of Practice Meeting

Panel VI: Ageing societies & Migration



**11,157** Staff members



**468** Locations



**128** Countries



**87%** in the field

**68.5 million** forcibly displaced people worldwide

Advocacy

Asylum and migration

Cash based interventions

Coordinating assistance

Education

Ending statelessness

Environment, disasters and climate change

Innovation

Livelihoods

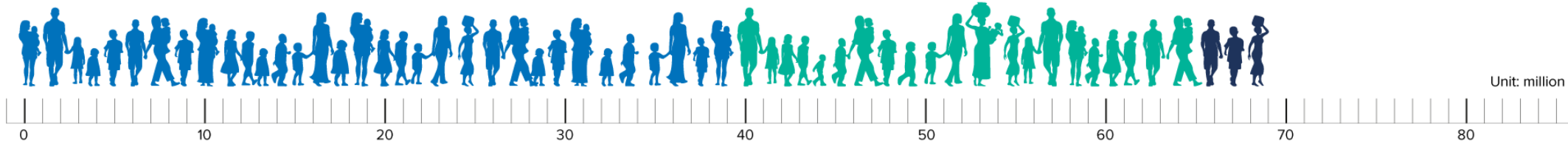
Protection

Public health

Safeguarding individuals

Shelter

Solutions



Internally Displaced People  
**40 million**

Refugees  
**25.4 million**

19.9 million under UNHCR mandate  
5.4 million Palestinian refugees registered by UNRWA

Asylum-seekers  
**3.1 million**

# *Context*

How to define & measure  
protection vulnerability

# *Options*

Structural Equation Model  
Logistic Regression  
Item Response Theory

# *Solution*

An R package to support  
implementation

# Context & problem statement

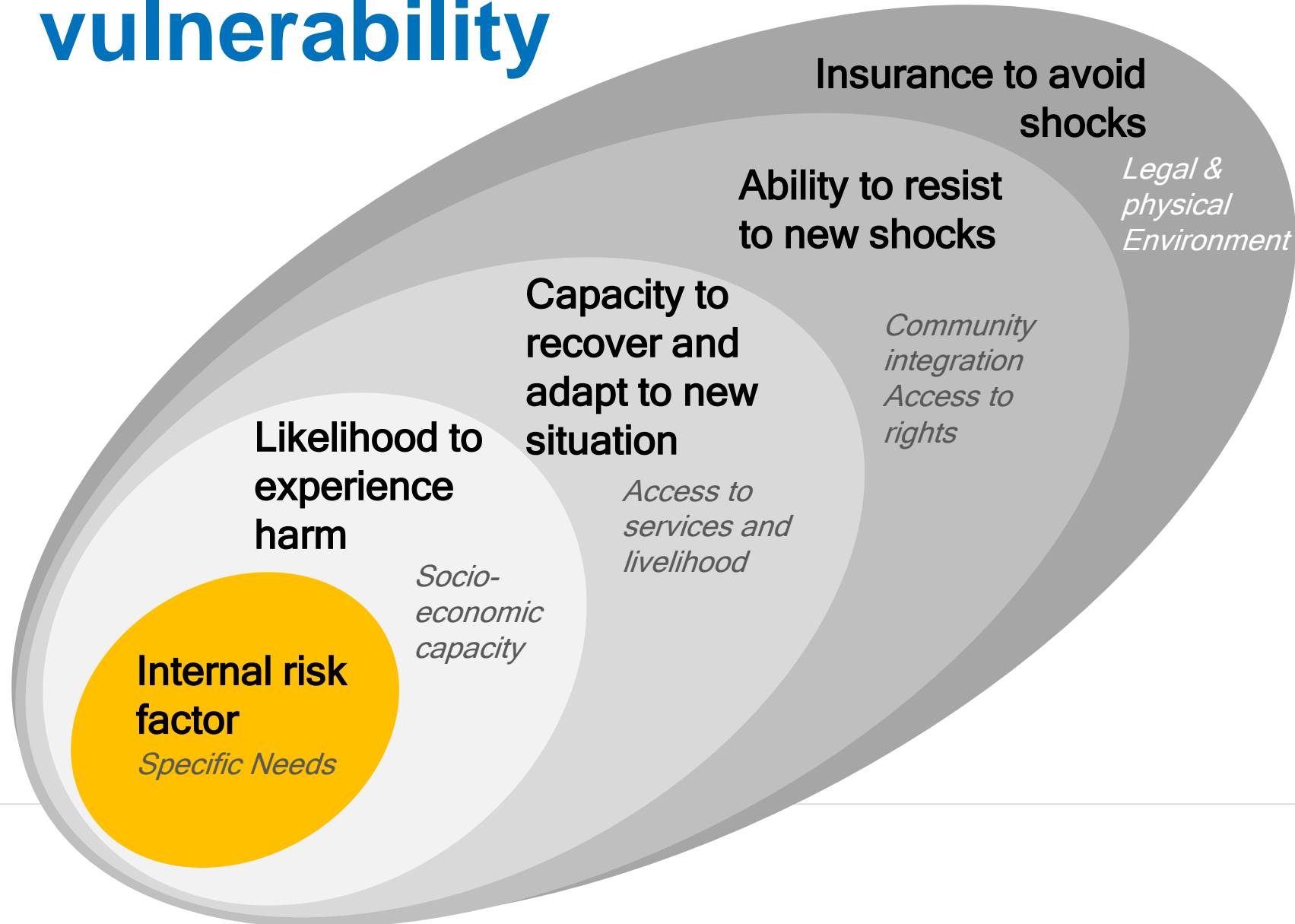
- Each operation in the region has a different approach to measure protection “vulnerability” in order to **prioritize and target assistance**
- **Harmonized** (rather than standardized) **approach for the measurement of vulnerability** is necessary

# If you can't measure it you can't improve it

$$\text{RISK} = \frac{\text{THREATS} \times \text{VULNERABILITIES}}{\text{CAPACITIES}}$$



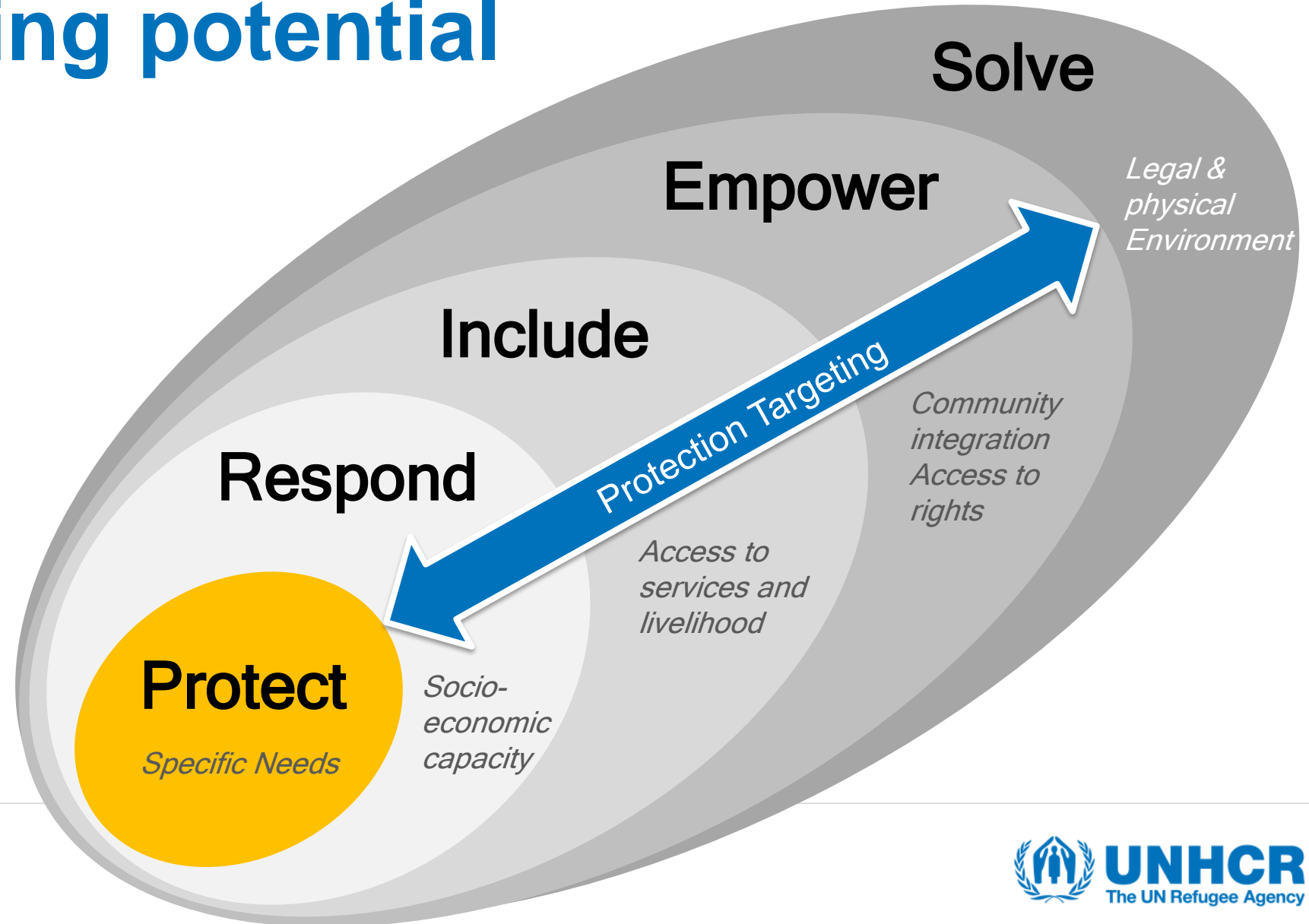
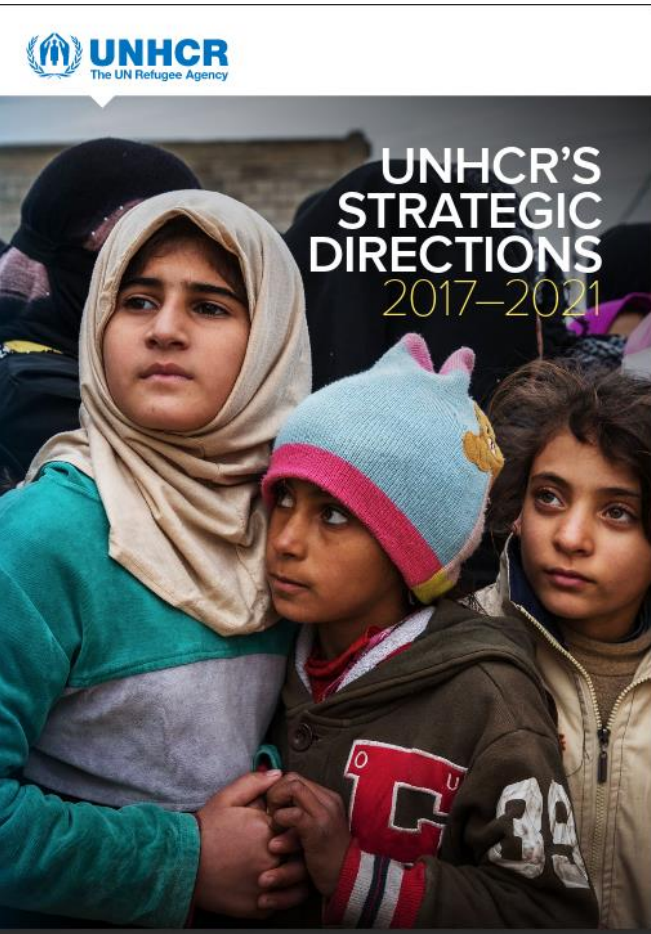
# Various connected concepts of vulnerability



Different types of vulnerability addressed by **multiple interventions**

Does not imply that there is a **sequence** between each type

# 5 Core Strategic Directions but only 3 implying potential targeting



# ***Protection Vulnerability Framework***

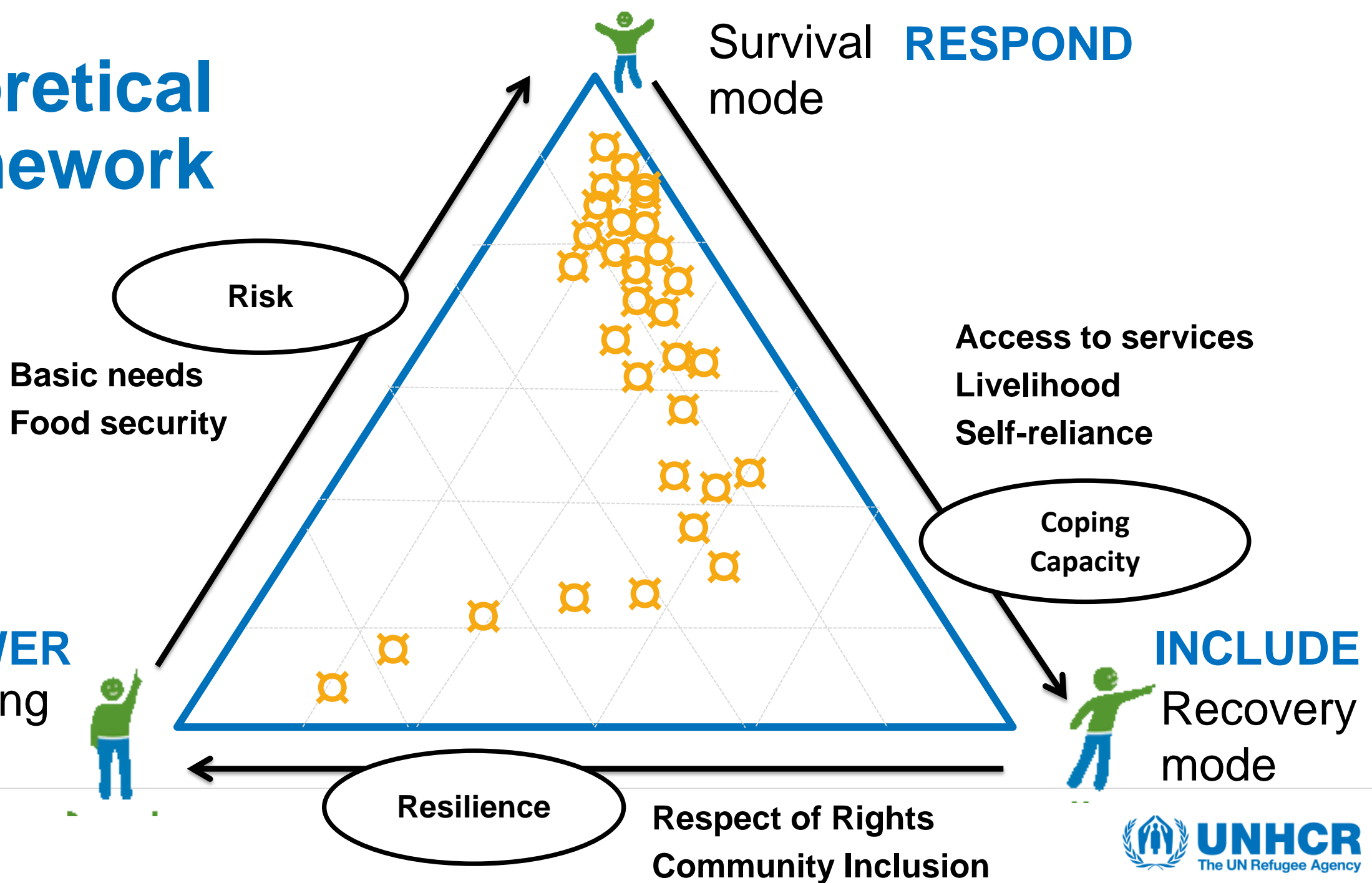
## Definition:

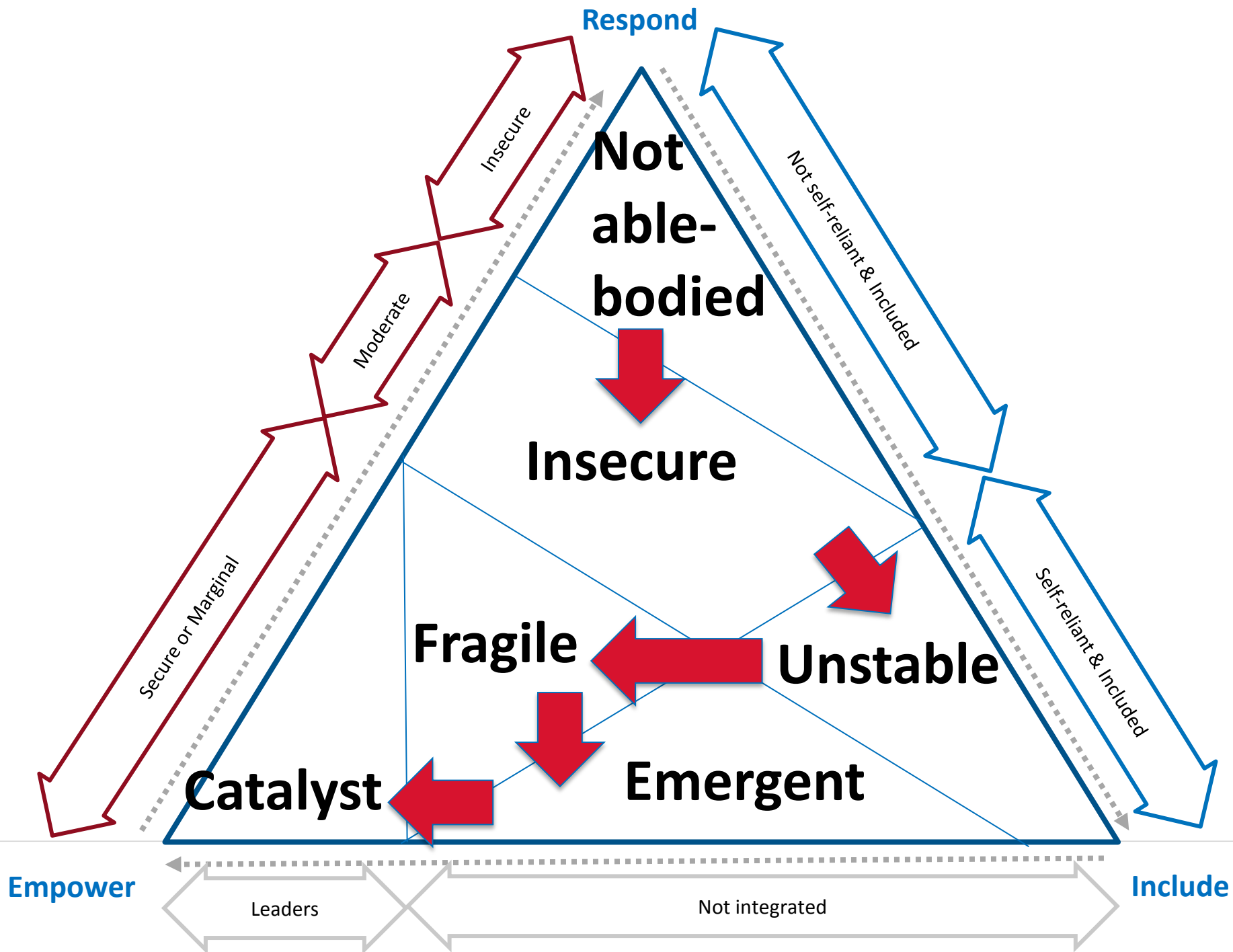
A framework to measure **household ability** to address different *dimensions* of vulnerability.

This framework aims at informing household **targeting needs** by articulating a series of potential assistance varying from In-kind distribution, Cash allowance, Livelihood support and Community building activities



# Theoretical Framework





# Building a *Protection Vulnerability Framework (PVF)* – A three-pronged approach

## 1. Theoretical framework with **three components**

Respond – Include – Empower

## 2. Conceptual Framework with three **measurable dimensions** (composite scores)

Risk – Coping – Resilience

## 3. Protection Scoring Tool

A statistical toolbox to help operations build scores and use results

# *Context*

How to define & measure  
protection vulnerability

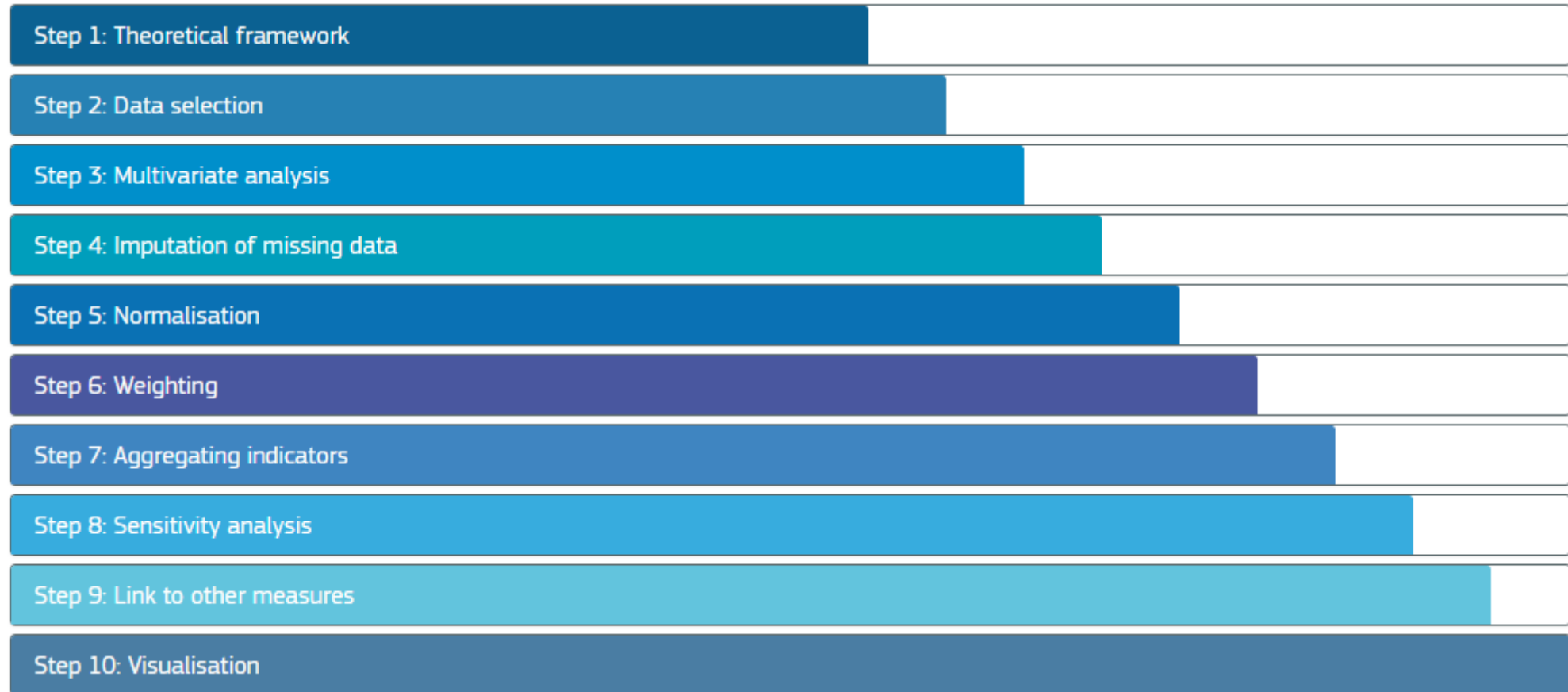
# *Options*

Factor analysis  
Structural Equation Model  
Item Response Theory

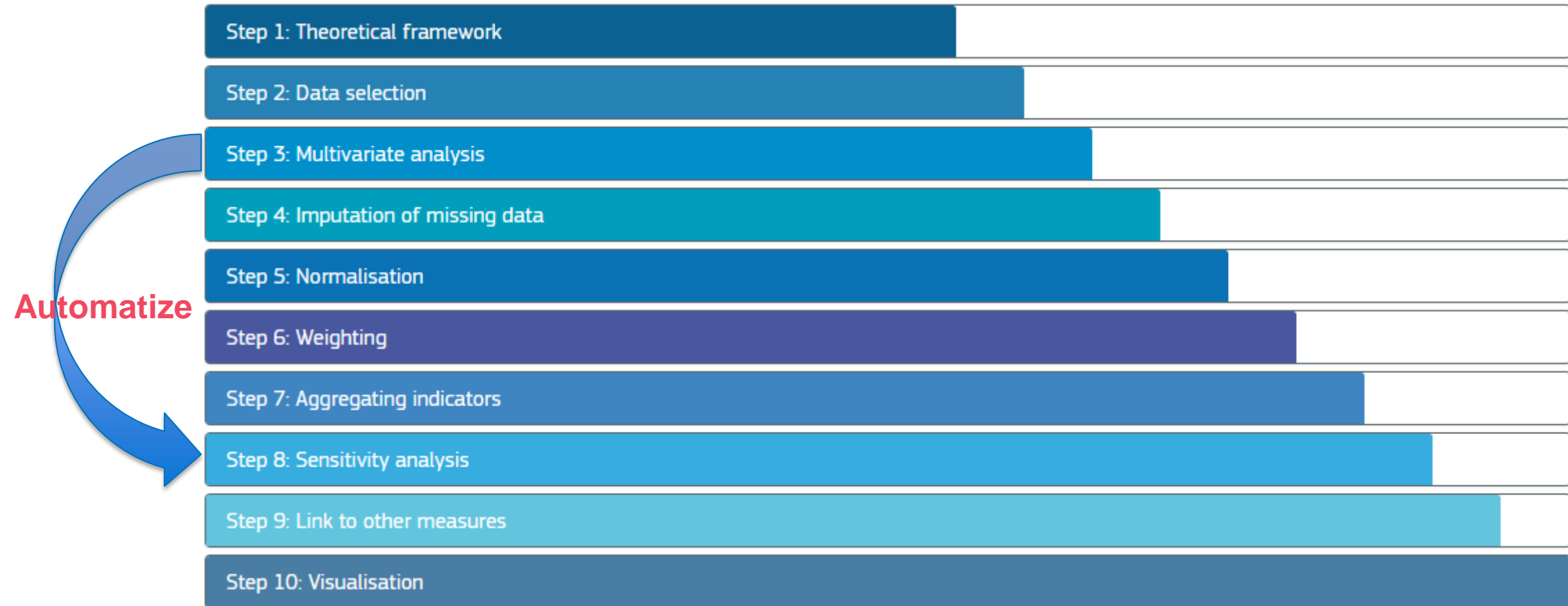
# *Solution*

An R package to support  
implementation

# OECD Handbook of Constructing Composite Indicator



# OECD Handbook of Constructing Composite Indicator



# Data set used for the research

- Sample based household Survey (2017) on Syrian Refugees in Kurdish Region of Iraq (“Joint Vulnerability Assessment”)
- Location: Refugee Camps and Urban settings
  - 2160 HHs
  - 450 Variables
  - 98 indicators used in the analysis (“experts”!!)

# Examples of Indicators

Eligible indicators are either:

- Based on administrative registry
- Declarative but potentially verifiable
- Indicators allocated to each dimension (index) by “experts!!”.

## Risk

- Main sources of drinking water
- Experienced lack of food in the past 30 days

## Coping

- Type of occupancy
- Benefited from primary health assistance

## Resilience

- Difficulties in obtaining birth certificates
- Children drop out from school



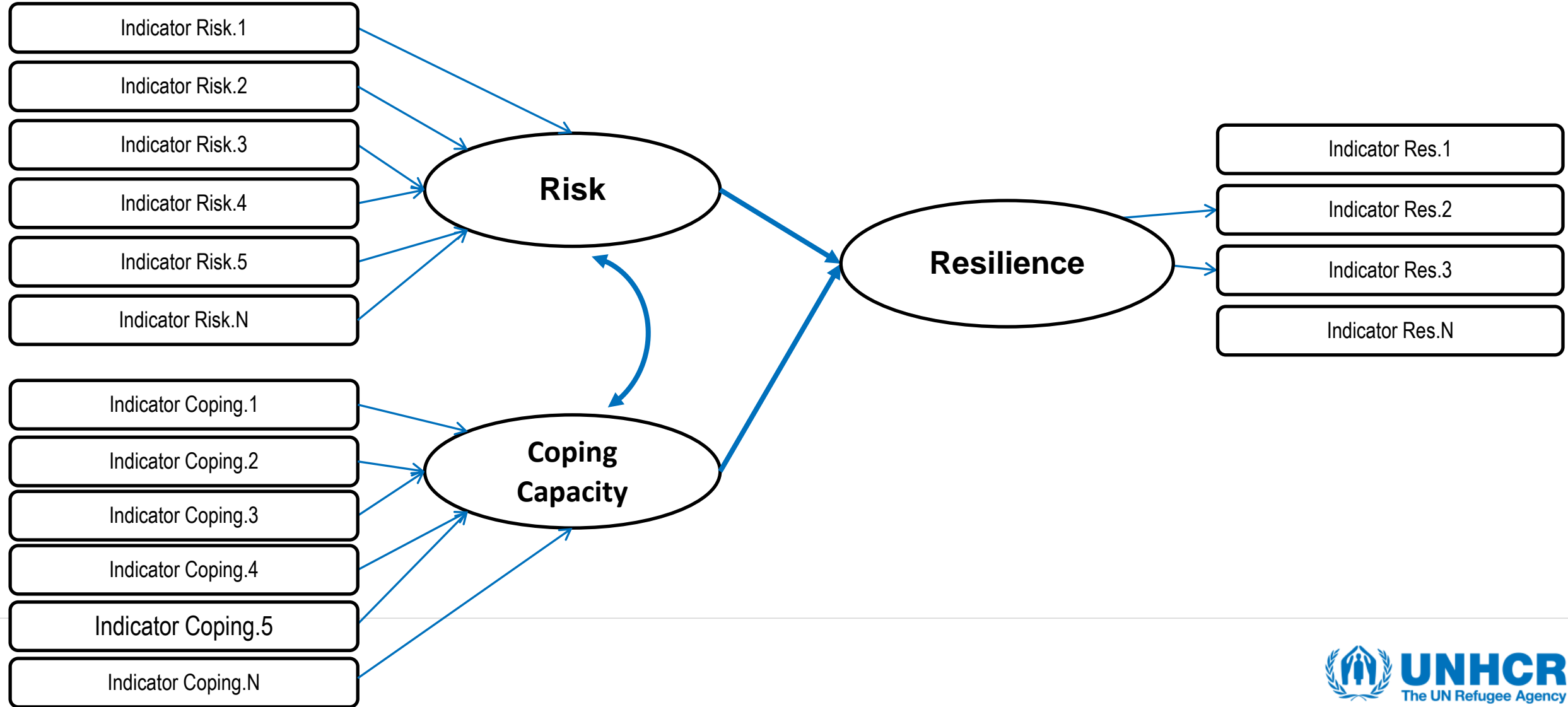
# Tested Methods

1. Factor Analysis (FA) and PCA
2. Structural Equation Modelling (SEM)
3. Item Response Theory (IRT)

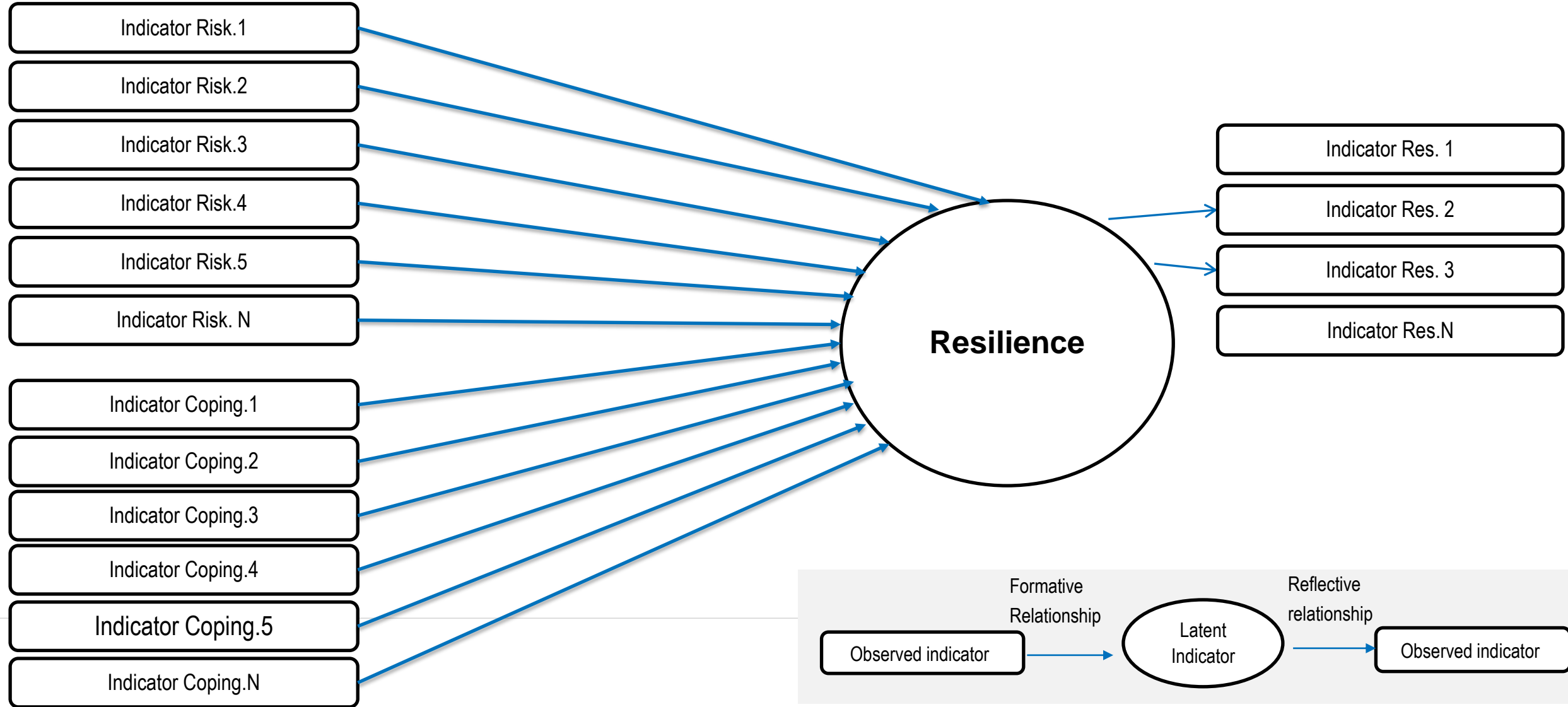
# Factor Analysis and SEM

- The factor analysis assumptions are severely violated
- The resulted factors and weights were questionable
- Equal weights did not work
- In SEM: high auto- and cross- correlation
- In SEM: The model reduced to linear regression

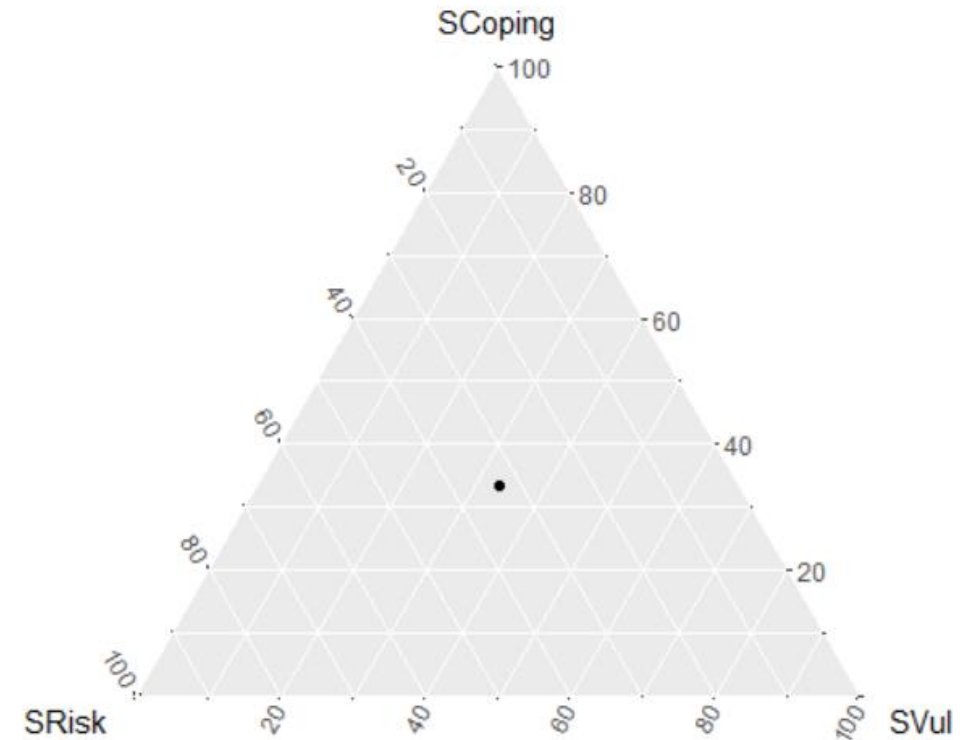
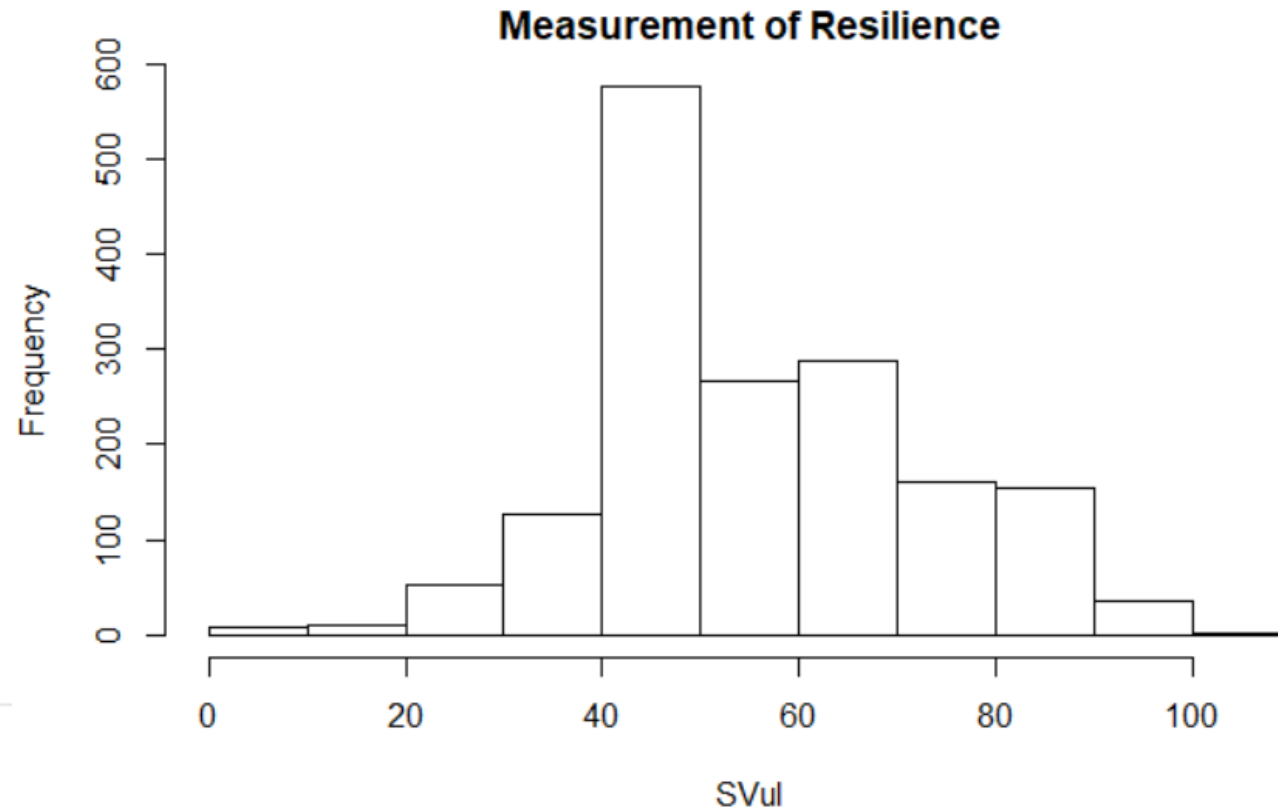
# SEM



# SEM (MIMIC)



# Problematic Results



# Main problems in phase 1

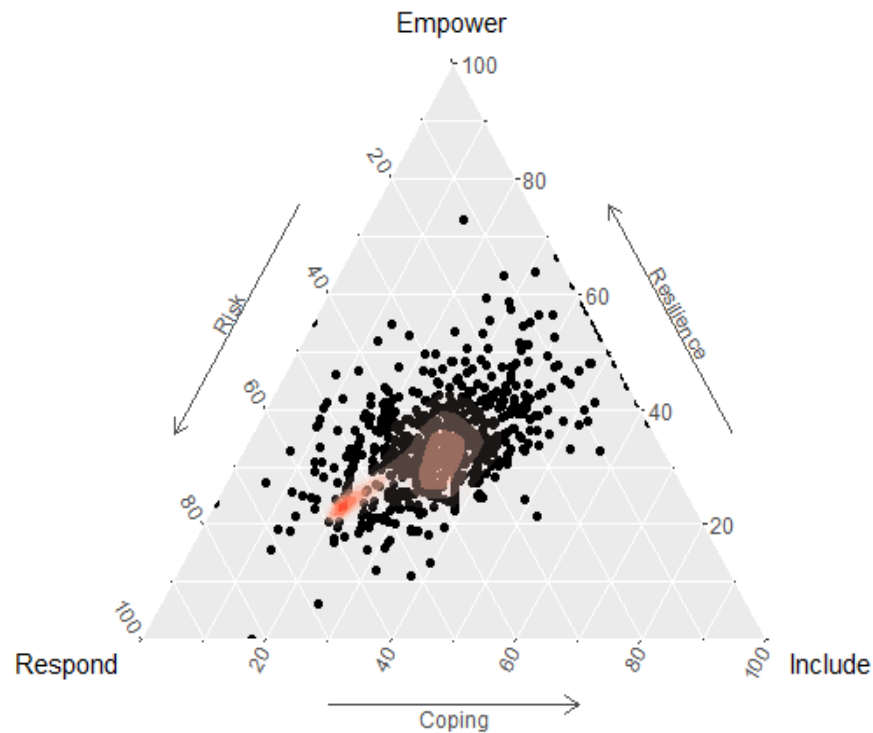
- All of the survey questions are categorical. Normalizing the data and scale is questionable!!
- The Indicators are highly correlated (auto-correlation)
- Cross loading issues (discriminant validity)
- Automatizing the process

# Changing the direction: Item Response Theory

		Observed indicators	
		<i>Continuous</i>	<i>Categorical</i>
Latent variables	<i>Continuous</i>	Factor analysis	Latent trait models
	<i>Categorical</i>	Latent profile analysis	Latent class models

# Latent Trait Model with equal weights

Protection Space for observed cases





# Remaining challenges

- Theoretically sound process to indicators inclusion and exclusion (selection criteria)
- Indicators weighting
- Automatizing the process

# Item Response Theory (IRT) in action

- Statistical tool for inferring an unobservable phenomenon (refugees' vulnerability), starting from a set of indicators (or variables)
- Extensions:
  - multidimensionality
  - discrete distribution for the underlying latent variable

# Multidimensionality

- Several, interrelated, dimensions contribute to define the general latent construct (vulnerability)
- **Exploratory** algorithm for clustering the variables → investigating the unknown dimensionality structure of the latent phenomenon

# Discrete distribution (latent class approach)

- Statistical units (i.e., refugees) divided into **several groups** on account of the (multidimensional) phenomenon (vulnerability)
- How many groups?
  - Subjective criteria (previous research, fixed by law, ...)
  - Objective criteria (statistical indexes for identifying the best number of groups)

# Weighting indicators

- Indicator weight = **discrimination parameter** from a multidimensional LC IRT model
- Discriminant power of an indicator = how much the indicator is able to distinguish units (refugees) with high latent trait (a dimension of vulnerability) w.r.t. units with low latent trait
- It is possible to weight the single indicators, taking into account the multidimensionality of the phenomenon under study

# Summarizing

1. Clustering of the indicators → **dimensionality assessment**
2. Clustering of the statistical units (refugees) → **refugees' profiling**
3. Weighting indicators → a **composite indicator** for each assessed dimension of vulnerability

# *Context*

How to define & measure  
protection vulnerability

# *Options*

Structural Equation Model  
Logistic Regression  
Item Response Theory

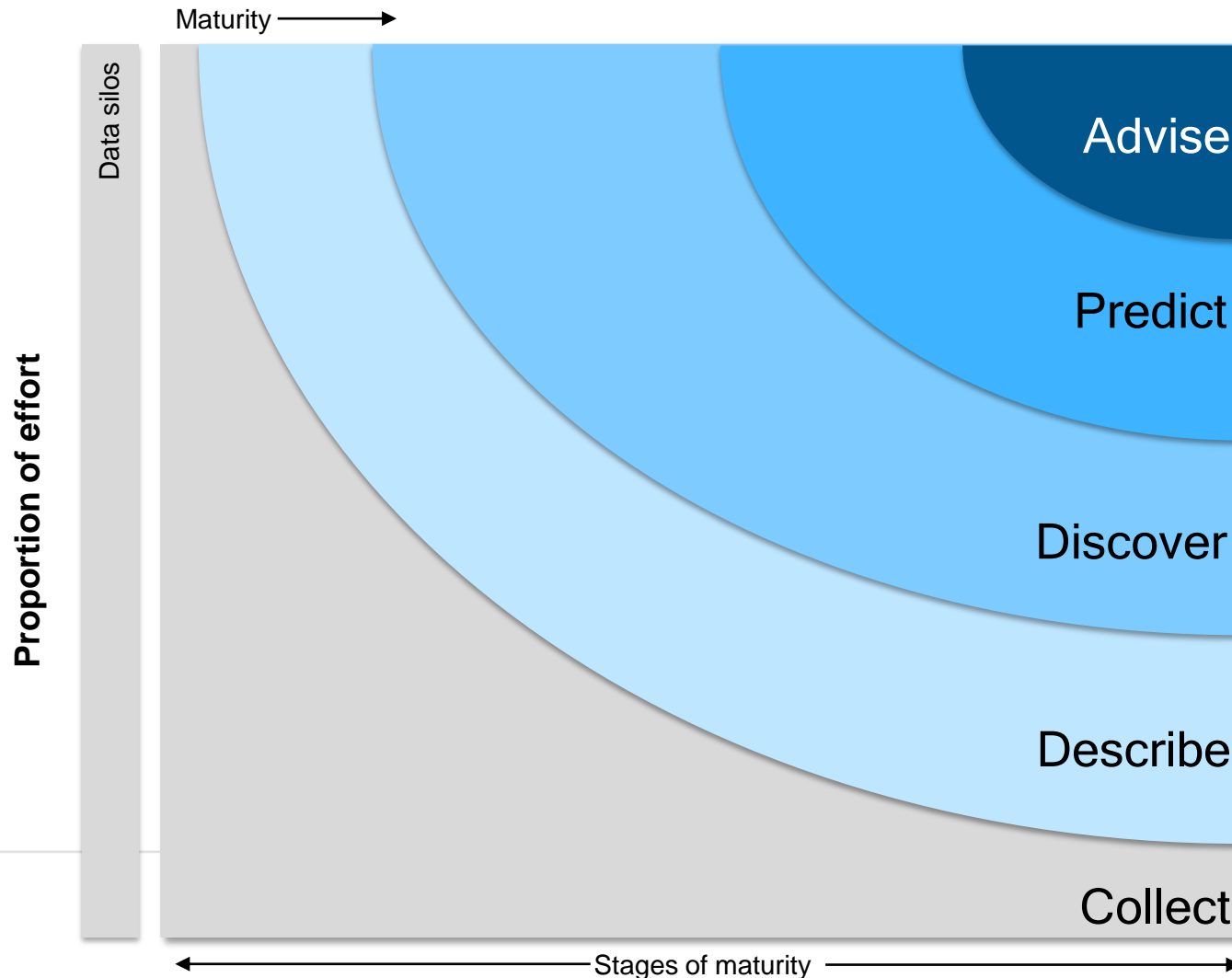
# *Solution*

An R package to support  
implementation

# Building the survey analytics toolbox with KoboLoad



The data science maturity model



**Vulnerability Multidimensional Scoring** for all monitored refugees

**Protection Risks prediction** for all registered Refugees

**Population segmentation** based on statistical profiles within survey

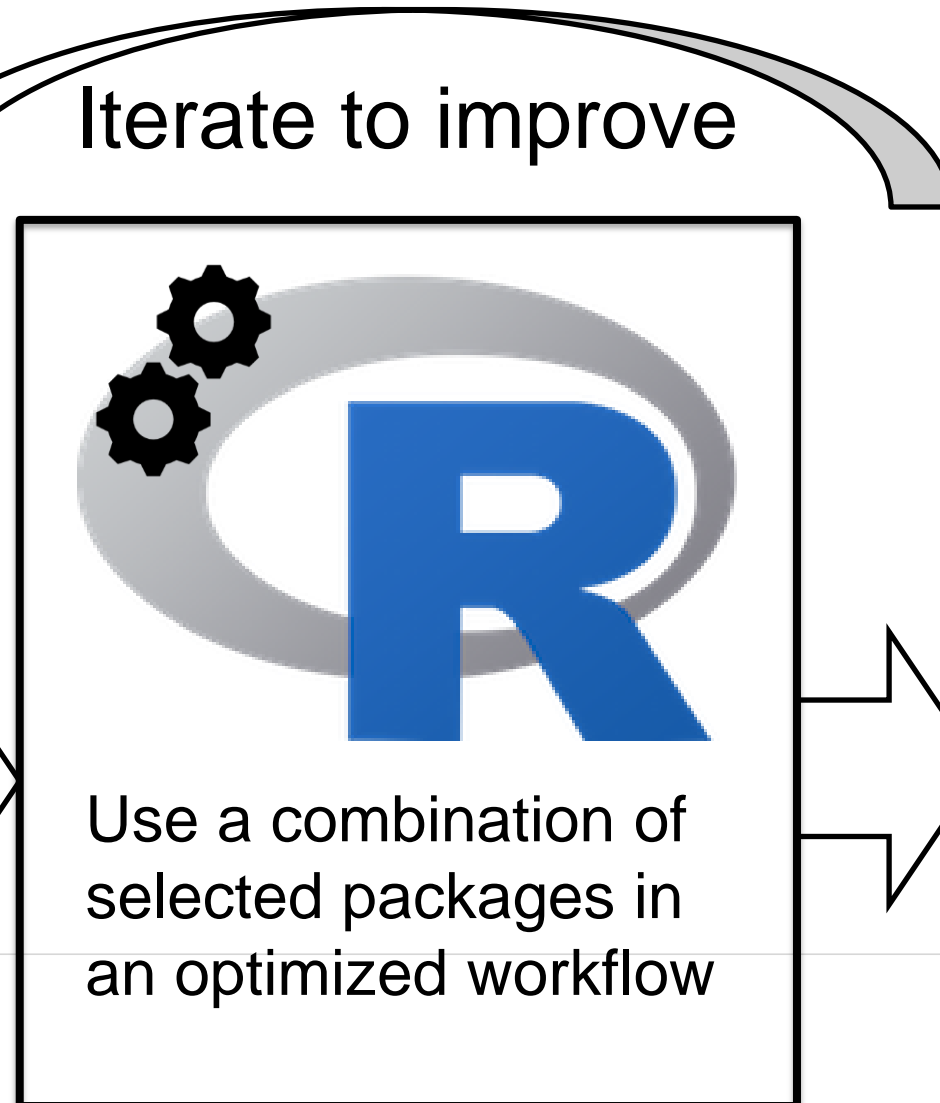
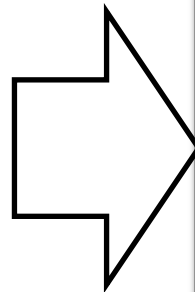
**Data insight** based on tabulations, crosstabulations, correlations, maps





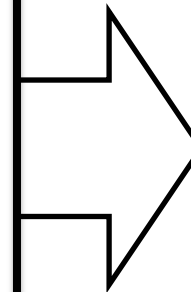
# KobolodeR approach: analysis iteration

**Configure the analysis plan** in Excel within the Xlsform used for the survey



Iterate to improve

Get standard report generated in Word



# Next...

- Confirm the soundness of the approach;
- Finalize and test the tool, train analyst in operations;
- Learn from experience coming from the implementation;
- Establish more research on most identified variables for vulnerability measurement.

# Contacts

- Edouard Legoupil – Head of the Data analysis Unit  
[Legoupil@unhcr.org](mailto:Legoupil@unhcr.org)
- Osama Abdelhay – Senior Data Scientist  
[Abdelhay@unhcr.org](mailto:Abdelhay@unhcr.org)
- Yara Maasri – Vulnerability Targeting Officer  
[Maasri@unhcr.org](mailto:Maasri@unhcr.org)
- Michela Gnaldi – Associate Professor at the University of Perugia  
[Michela.Gnaldi@unipg.it](mailto:Michela.Gnaldi@unipg.it)
- Simone Del Sarto – Postgraduate researcher at the University of Perugia  
[Simone.delsarto@email.com](mailto:Simone.delsarto@email.com)