Building a Protection Vulnerability Formula

JRC Community of Practice Meeting

Panel VI: Ageing societies & Migration
11,157 Staff members in 468 Locations in 128 Countries

87% in the field

68.5 million forcibly displaced people worldwide

Internally Displaced People 40 million
Refugees 25.4 million
Asylum-seekers 3.1 million

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
**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
Various connected concepts of vulnerability

Different types of vulnerability addressed by multiple interventions

Does not imply that there is a sequence between each type

Internal risk factor

*Specific Needs*

- **Likelihood to experience harm**
- **Socio-economic capacity**
- **Ability to resist to new shocks**
- **Capacity to recover and adapt to new situation**
- **Access to services and livelihood**
- **Community integration Access to rights**
- **Insurance to avoid shocks**
- **Legal & physical Environment**
5 Core Strategic Directions but only 3 implying potential targeting

Respond

Include

Empower

Solve

Protect

Socio-economic capacity

Specific Needs

Protection Targeting

Community integration Access to rights

Access to services and livelihood

Legal & physical Environment
**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

- **RESPOND**
  - Survival mode
  - Risk
  - Basic needs (Food security)
  - Access to services
  - Coping Capacity

- **EMPOWER**
  - Supporting mode
  - Resilience
  - Respect of Rights
  - Community Inclusion

- **INCLUDE**
  - Recovery mode
  - Basic needs (Livelihood, Self-reliance)

- **Theoretical Framework**
  - UNHCR
  - The UN Refugee Agency
Not able-bodied

Insecure

Fragile

Unstable

Emergent

Respond

Include

Empower

Secure or Marginal

Moderate

Not self-reliant & included

Self-reliant & included

Catalyst

Not integrated

Leaders
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

Step 1: Theoretical framework
Step 2: Data selection
Step 3: Multivariate analysis
Step 4: Imputation of missing data
Step 5: Normalisation
Step 6: Weighting
Step 7: Aggregating indicators
Step 8: Sensitivity analysis
Step 9: Link to other measures
Step 10: Visualisation
OECD Handbook of Constructing Composite Indicator

- Step 1: Theoretical framework
- Step 2: Data selection
- Step 3: Multivariate analysis
- Step 4: Imputation of missing data
- Step 5: Normalisation
- Step 6: Weighting
- Step 7: Aggregating indicators
- Step 8: Sensitivity analysis
- Step 9: Link to other measures
- Step 10: Visualisation

Automatize
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!!”.

<table>
<thead>
<tr>
<th>Risk</th>
<th>Coping</th>
<th>Resilience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main sources of drinking water</td>
<td>Type of occupancy</td>
<td>Difficulties in obtaining birth certificates</td>
</tr>
<tr>
<td>Experienced lack of food in the past 30 days</td>
<td>Benefited from primary health assistance</td>
<td>Children drop out from school</td>
</tr>
</tbody>
</table>
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
Problematic Results

Measurement of Resilience

UNHCR
The UN Refugee Agency
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

<table>
<thead>
<tr>
<th>Latent variables</th>
<th>Observed indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Continuous</strong></td>
<td>Continuous</td>
</tr>
<tr>
<td><strong>Categorical</strong></td>
<td>Factor analysis</td>
</tr>
<tr>
<td></td>
<td>Latent profile analysis</td>
</tr>
<tr>
<td></td>
<td>Latent trait models</td>
</tr>
<tr>
<td></td>
<td>Latent class models</td>
</tr>
</tbody>
</table>
Latent Trait Model with equal weights
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 KoboLoade

The data science maturity model

Proportion of effort

Maturity

Data silos

Collect

Describe

Discover

Predict

Advise

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
KoboloadeR 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

Use a combination of selected packages in an optimized workflow

[https://unhcr.github.io/koboloadeR/docs/](https://unhcr.github.io/koboloadeR/docs/)
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 or the Data analysis Unit
  Legoupil@unhcr.org
• Osama Abdelhay – Senior Data Scientist
  Abdelhay@unhcr.org
• Yara Maasri – Vulnerability Targeting Officer
  Maasri@unhcr.org
• Michela Gnaldi – Associate Professor at the University of Perugia
  Michela.Gnaldi@unipg.it
• Simone Del Sarto – Postgraduate researcher at the University of Perugia
  Simone.delsarto@email.com