Groundwork to produce nowcasts on Intentional Homicide
Nowcasts, why are they needed

Data timeliness key requirement for:

• monitoring sustainable development

• providing timely evidence for policy making

• particularly important for certain topics (violence, drug trafficking routes, spread of psychoactive substances)
Priorities for nowcasts production

- **Violent crime**
  - 16.1.1, Number of victims of intentional homicide per 100,000 population, by sex (victims per 100,000 population)

- **Criminal Justice System**
  - 16.3.2, Unsentenced detainees as a proportion of overall prison population (%)

- **Drug use**
  - Prevalence of drug use
  - Drug-related mortality

- **Trafficking routes based on individual seizures data (drugs, protected wildlife, firearms)**
### UNODC data collections

**Drugs:**
- Demand (use, treatment, mortality, etc.)
- Supply (production, seizures, prices, etc.)
- Monitoring of new substances

**Crime:**
- Violent crime and other crimes
- Criminal justice operations
- Trafficking in persons
- Trafficking of firearms
- Trafficking of protected wildlife

Data sources: officially collected data from Member States, complemented with data from other International Organizations, academia, NGOs
Challenges to produce timely data on crime and drugs:

- Current data collection/production cycle results in a 2-year gap between reference and publication year.
- Data availability still a challenge for certain regions (e.g. Africa) and certain topics (e.g. drug use).
- Data standardisation/comparability still developing.
- With some exceptions, long time series often not available.
- Political sensitivity of several data topics: use of national official data is preferred, use of other data needs to be explained.
Case study: applying the **Exponential Weighted Moving Average** to Homicide data

Why the Exponential Weighted Moving average?

- Uses all available information to impute missing values as weighted average of the last \( n \) available data-points, where the weight decreases exponentially with each previous period. Any imputed value receives higher contribution from the more recently available information.

- Produces “conservative” imputations, homicide rates are generally stable in time

- Provides an adaptive algorithm that can be used for all countries, irrespective of homicide patterns and differences in data availability
Why have we started from the Homicide study?

Country coverage (based on 2017 official homicide rates)
Exponential Weighted Moving Average: results

variable: actual, imputed

France

Germany

Greece
Exponential Weighted Moving Average: results

variable: actual, imputed
Where does the model fit the best?

<table>
<thead>
<tr>
<th>country name</th>
<th>year</th>
<th>actual</th>
<th>imputed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>2018</td>
<td>6.6555472</td>
<td>7.0224903</td>
</tr>
<tr>
<td>Argentina</td>
<td>2018</td>
<td>5.3244967</td>
<td>5.7577912</td>
</tr>
<tr>
<td>Australia</td>
<td>2018</td>
<td>0.8916379</td>
<td>0.9013716</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>2018</td>
<td>2.2010050</td>
<td>2.1441501</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>2018</td>
<td>2.3733246</td>
<td>2.3597637</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>2018</td>
<td>1.4747589</td>
<td>1.4233573</td>
</tr>
<tr>
<td>Colombia</td>
<td>2018</td>
<td>25.5069370</td>
<td>25.8038220</td>
</tr>
<tr>
<td>Germany</td>
<td>2018</td>
<td>0.9479813</td>
<td>0.9975694</td>
</tr>
<tr>
<td>Greece</td>
<td>2018</td>
<td>0.9408858</td>
<td>0.8505299</td>
</tr>
<tr>
<td>Ireland</td>
<td>2018</td>
<td>0.8715501</td>
<td>0.8507856</td>
</tr>
<tr>
<td>Mongolia</td>
<td>2018</td>
<td>6.1829653</td>
<td>6.1632965</td>
</tr>
<tr>
<td>Poland</td>
<td>2018</td>
<td>0.7304467</td>
<td>0.7526222</td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>2018</td>
<td>0.6038459</td>
<td>0.6597254</td>
</tr>
<tr>
<td>Republic of Moldova</td>
<td>2018</td>
<td>3.8993090</td>
<td>4.0703482</td>
</tr>
<tr>
<td>Serbia</td>
<td>2018</td>
<td>1.2268545</td>
<td>1.2034826</td>
</tr>
<tr>
<td>South Africa</td>
<td>2018</td>
<td>36.3988718</td>
<td>34.5112462</td>
</tr>
<tr>
<td>Spain</td>
<td>2018</td>
<td>0.6210781</td>
<td>0.6563051</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>2018</td>
<td>2.4212163</td>
<td>2.4513461</td>
</tr>
<tr>
<td>Sweden</td>
<td>2018</td>
<td>1.0830325</td>
<td>1.0930806</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2018</td>
<td>0.5864415</td>
<td>0.5580909</td>
</tr>
<tr>
<td>Uganda</td>
<td>2018</td>
<td>10.5244682</td>
<td>11.4508054</td>
</tr>
</tbody>
</table>
Country name                         year  actual    imputed
Armenia                             2018  1.69376694  2.5465437
Croatia                             2018  0.57747834  1.0452631
Kenya                               2018  3.05294495  4.8841700
Kyrgyzstan                          2018  2.18908629  4.5814374
Oman                                2018  0.26920688  0.5418518
Singapore                           2018  0.15633142  0.2415968
Slovenia                            2018  0.48123195  0.7877614
State of Palestine                 2018  0.39070533  0.7811142
Venezuela (Bolivarian Republic of) 2018 36.07851283 57.0686400

Where does the model fail?
Exponential Weighted Moving Average: conclusion

**Pros**
- The exponential EWMA provides an adaptive algorithm which can be used both to impute missing values (fill the gaps in the time-series) and to extrapolate/nowcast homicide rates
- It uses all available information in a time series
- Broad availability of R open-source packages (imputeTS)

**Cons**
- Extrapolated values always lay in the range between the min and max of the time series
- The model fails in predicting extraordinary events
Nowcast and new technologies

Explore ‘early warning’ information through technologies based on **web scraping**

Future initiatives:

- Test methodology to web scrape media news on specific events (drug seizures and homicides) to identify and reports on them
- Use the potential of news reported on media and social media to **complement available official data**

Challenges:

- Properly identify single events, develop algorithms to clean duplicates
- Manage multi-source databases and embed this new data source in the methodological framework to produce nowcast.
Thank you!