Annex A. Data and methodology

The data

The main producing economies of fake pharmaceuticals and the key transit points are determined using statistical "filters" (see below). This is done based on three sources of information:

- data on seizures of counterfeit pharmaceuticals.
- international trade statistics on the pharmaceutical sector, and
- industrial activity data for the pharmaceutical sector.

An important data limitation should be highlighted in this context. While the quality of data on customs seizures of infringing pharmaceutical products received from member countries of the EU and from the US is very high, the data from South American, African, Middle Eastern and Asian customs authorities are of insufficient quality. Hence the mapping exercise for the EU and the US as destinations is relatively precise, but a precise charting of trade routes and the modes of transport for the other regions is not possible. For transparency purposes, all data gaps were highlighted throughout the analysis.

In addition, the datasets identify a set of EU member countries as provenances. However, these identifications are based on data from the European Commission's Directorate-General for Taxation and Customs Union (DG TAXUD), and refer to goods coming from outside the EU that were seized in a different member state to the country where it entered the EU. This is because DG TAXUD data refer only to imports to the EU from third countries, and do not include internal EU trade.

Data on seizures of counterfeit goods

The database on customs seizures is the critical quantitative input to this study. It was constructed from three separate datasets received from the WCO, from DG TAXUD of the European Commission, and from the US Department of Homeland Security. The database includes detailed information on seizures of IPR-infringing goods made by customs officers in 99 economies around the world between 2014 and 2016. For each year, there are more than 100 000 observations in the database; in most cases one observation corresponds to one customs seizure.

The database contains a wealth of information about IPR-infringing goods that can be used for quantitative and qualitative analysis. In most cases the database reports for each seizure: date of seizure, mode of transport of fake products, departure and destination economies, general statistical category of seized goods as well as their detailed description, name of legitimate brand owner, number of seized products and their approximate value.

Concerning valuation of seized goods, there are two principles for reporting the value of counterfeit goods: 1) declared value (value indicated on customs declarations), which corresponds to values reported in the general trade statistics; and 2) replacement value (price of original goods). The structured interviews with customs officials and the descriptive analysis of values of selected products conducted in OECD/EUIPO (2016) revealed that the declared values are reported in most cases.

International trade statistics

The trade statistics are based on the United Nations (UN) Comtrade database (landed customs value). With 171 reporting economies and 247 partner economies (76 economies in addition to reporting economies), the database covers the largest part of world trade and is considered the most comprehensive trade database available. Products are registered on a six-digit Harmonized System (HS) basis (see WCO, 2019), and can then be aggregated.

This study uses two different types of trade statistics provided by the UN Comtrade database. First, the calculations of the General Trade Related Indices (GTRIC) are based on import data. Second, the identification of potential transit points is based on re-export data. Re-exports are exports of foreign goods in the same state as previously imported, i.e., that have not acquired domestic origin through processing.

In most economies, import statistics are compiled from the records filed with local customs authorities. This is particularly important in the context of this report as data on customs seizures of infringing products originate from the same source – customs offices at the destination. This reinforces the choice for import statistics as the reference point for the calculation of the GTRIC indices, as both imports data and seizure data refer to the same observed incoming trade flows.

Industrial activity data

The identification of potential producer points of fake pharmaceutical goods and medicines is based on data on industrial activity provided by the UNIDO Industrial Statistics Database (UNIDO, 2019). This study takes advantage of the cross-country comparability of the data on industrial output and value-added included in the UNIDO's Industrial Statistics Activity database (UNIDO, 2019) to distinguish a producing economy from a potential transit point for the pharmaceutical sector. The database contains seven principal indicators of industrial statistics (number of establishments, number of employees, wages and salaries, output, value added, gross fixed capital formation, and number of female employees) at the 4-digit level of the International Standard Industrial Classification of All Economic Activities (ISIC).

The main producing economies and key transit points for counterfeit pharmaceuticals were identified following several steps:

1. Economies were ranked according to their propensity to be an economy of provenance for counterfeit pharmaceuticals. The resulting index is called GTRIC-e. These indices are calculated in Chapter 4, and economies more likely to export counterfeit pharmaceuticals are presented in Table 4.1.
2. An indicator of the relative comparative advantage for producing pharmaceuticals was calculated for each economy (RCAP-e) based on UNIDO (2019) data. This is the first "filter" to be used in the analysis. The methodology is described in the next subsection of this annex.
3. For each economy an indicator of the relative comparative advantage for being a transit point in global trade in pharmaceuticals was calculated (RCAT-e) based on re-export data (UN Trade Statistics Division, 2019). This is the second "filter" to be used in the analysis. The methodology is described in the next subsection of this annex.
4. Both filters (RCAP-e and RCAT-e indicators) were applied for every economy with a high GTRIC-e score. This indicates whether the given economy is a producing one, or a potential transit point for fake pharmaceuticals.
5. Some additional descriptive statistical analysis checked the modes of transport and the size of shipments on the selected trade routes.

It should be highlighted that the framework presented below relies on a set of methodological assumptions. For transparency purposes all are spelt out in the text.

Construction of GTRIC-e for pharmaceuticals

The first step was to rank all the known provenance economies by their relative intensity of exporting fake pharmaceuticals. This distinguished the key provenances in trade with fake pharmaceuticals. Each of these key points then was investigated further to determine its exact role in trade in fake pharmaceutical products and medicines.

The most intense provenance economies were identified using an index that ranked them according to their relative propensity to be an economy of provenance for counterfeit pharmaceuticals (GTRIC-e). The index is based on the data on global customs seizures and data on imports (OECD/EUIPO, 2019). It takes into account: 1) the absolute value of exports of fake pharmaceuticals from a given economy (in USD); and 2) the share of fakes in total exports of fake pharmaceuticals from a given economy.

The construction of GTRIC-e directly relied on the methodology introduced in the OECD/EUIPO (2019) study. A detailed description of the methodology used to calculate the GTRIC-e is provided below.

Importantly, two assumptions are made to calculate the GTRIC vectors. The first is that the volume of seizures of a given product or from a given source economy is positively correlated with the actual intensity of trade in counterfeit goods in that product category or from that economy. The second assumption acknowledges that this relationship is not linear, as there might be some biases in the detection and seizure procedures. For instance, the fact that infringing goods are detected more frequently in certain categories could imply that differences in counterfeiting factors across products merely reflect that some goods are easier to detect than others, or that some goods, for one reason or another, have been specially targeted for inspection.

GTRIC-e was constructed in four steps:

1. For each reporting economy, the seizure percentages for provenance economies were calculated.
2. For each provenance economy, aggregate seizure percentages were formed, taking the reporting economies' share of sensitive imports as weights.
3. From these, each economy's counterfeit source factor was established, based on the provenance economies' weight in terms of global trade.
4. Based on these factors, the GTRIC-e was formed.

Step 1: Measuring reporter-specific seizure intensities from each provenance economy

v_{epi} is economy *i*'s registered seizures of all types of infringing goods included in a given product category *p* that originate from economy *e* at a given year in terms of value.

γ_{epi} is economy *i*'s relative seizure intensity (seizure percentage) of all infringing items within the product category that originate from economy *e*, in a given year:

$$\gamma_{epi} = \frac{v_{epi}}{\sum_e v_{epi}}, \text{ such that } \sum_e \gamma_{epi} = 1 \ \ \forall i$$

Step 2: Measuring general seizure intensities of each provenance economy

The general seizure intensity for economy *e* within the product category *p*, denoted Γ_{ep}, is then determined by averaging seizure intensities, γ_{epi}, weighted by the reporting economy's share of world imports from known counterfeit and pirate origins.[1] Hence:

$$\Gamma_{ep} = \sum_i \varpi_{pi} \gamma_{epi}$$

where the weight of reporting economy *i* is given by

$$\varpi_{pi} = \frac{m_{epi}}{\sum_i m_{epi}}$$

with m_{epi} is economy *i*'s imports of goods in a given product category *p* from economy *e* at a given year in terms of value, so that $\sum_i \varpi_{pi} = 1 \ \forall p$

Step 3: Measuring partner-specific counterfeiting factors

$m_{ep} = \sum_i m_{epi}$ is defined as the total registered world imports of all sensitive goods in the product category *p* from provenance economy *e*.

$m_p = \sum_e m_{ep}$ is defined as the total registered world imports of all sensitive goods in the product category *p* from all provenance economies.

The share of provenance economy *e* in world imports of all sensitive goods in the product category *p*, denoted s_{ep}, is then given by:

$$s_{ep} = \frac{m_{ep}}{m_p}, \text{ such that } \sum_e s_{ep} = 1, \ \forall p$$

From this, the economy-specific counterfeiting factor is established by dividing the general seizure intensity for economy *e* with the share of world imports from *e* within the product category *p*:

Step 4: Establishing GTRIC-e

Gauging the magnitude of counterfeiting and piracy from a provenance economy perspective can be done in a similar fashion as for sensitive goods. Hence, a general trade-related index of counterfeiting for economies (GTRIC-e) is established along similar lines and assumptions:

- The first assumption (A3) is that the intensity by which any counterfeit or pirated article from a particular economy is detected and seized by customs is positively correlated with the actual amount of counterfeit and pirate articles imported from that location.
- The second assumption (A4) acknowledges that assumption A3 may not be entirely correct. For instance, a high seizure intensity of counterfeit or pirated articles from a particular provenance economy could be an indication that the provenance economy is part of a customs profiling scheme, or that it is specially targeted for investigation by customs. The importance that provenance economies with low seizure intensities play regarding actual counterfeiting and piracy activity could therefore be under-represented by the index and lead to an underestimation of the scale of counterfeiting and piracy.

As with the product-specific index, GTRIC-e is established by applying a positive monotonic transformation of the counterfeiting factor index for provenance economies using natural logarithms. This follows from assumption A3 (positive correlation between seizure intensities and actual infringement activities) and assumption A4 (lower intensities tend to underestimate actual activities). Considering the

possibilities of outliers at both ends of the GTRIC-e distribution – i.e. some economies may be wrongly measured as being particularly susceptible sources of counterfeit and pirated imports, and vice versa – GTRIC-e is approximated by a left-truncated normal distribution as it does not take values below zero.

The transformed general counterfeiting factor across provenance economies on which GTRIC-e is based is therefore given by applying logarithms onto economy-specific general counterfeit factors (see, for example, Verbeek, 2000):

$$cf_{ep} = \ln(CF_{ep} + 1)$$

In addition, it is assumed that GTRIC-e follows a truncated normal distribution with $cf_{ep} \geq 0$. Following Hald (1952), the density function of the left-truncated normal distribution for cf_{ep} is given by

$$g_{LTN}(cf_{ep}) = \begin{cases} 0 & if\ cf_{ep} \leq 0 \\ \dfrac{g(cf_{ep})}{\int_0^{\infty} g(cf_{ep})\partial cf_{ep}} & if\ cf_{ep} \geq 0 \end{cases}$$

Where $g(cf_{ep})$ is the non-truncated normal distribution for cf_{ep} specified as:

$$g(cf_{ep}) = \frac{1}{\sqrt{2\pi\sigma_{cf}^2}} \exp\left(-\frac{1}{2}\left(\frac{cf_{ep} - \mu_{cf}}{\sigma_{cf}} \right)^2 \right)$$

The mean and variance of the normal distribution, here denoted μ_{cf} and σ_{cf}^2, are estimated over the transformed counterfeiting factor index, cf_{ep}, and given by $\widehat{\mu}_{cf}$ and $\widehat{\sigma}_{cf}^2$.

This enables the calculation of the counterfeit import propensity index within each product category *p* (GTRIC-e) across provenance economies, corresponding to the cumulative distribution function of cf_{ep}.

Methodology to identify producers from transit points of counterfeit pharmaceuticals

Construction of RCAP-e and RCAT-e

Relative comparative advantage for production of a given good (RCAP-e)

The first statistical filter that can be used to tell producers from transit points looks at the production capacities of a given economy in the pharmaceutical sector. The rationale behind this test is simple: production activity often relies on certain skills, or resources. It also exhibits certain returns to scale properties that results in specialisation of this particular economy in the production of pharmaceuticals. Hence, production of counterfeit medicines and pharmaceutical goods is more likely to occur in a known provenance economy that specialises in the legitimate production of pharmaceuticals, than in a country without production capacity in the pharmaceutical sector.

This specialisation of a given trading economy in production of pharmaceuticals is captured by an indicator of the relative comparative advantage for production (RCAP-e). The indicator looks at the share of industrial activity in the pharmaceutical sector with the total industrial activity in a given economy.

Construction of this indicator is based on industry statistics. Importantly, these statistics are based on a different taxonomy than the trade statistics, hence a matching exercise was performed (see Box 5). A detailed description of the methodology used to calculate the RCAP-e is provided below.

Formally, the revealed comparative advantage in production for an economy *e* in the pharmaceutical sector (RCAP) measures whether this economy produces more pharmaceuticals as a share of its total production than the "average" country:

$$RCAP_{ep} = \frac{y_{ep} \Big/ \sum_p y_{ep}}{\sum_e y_{ep} \Big/ \sum_e \sum_p y_{ep}}$$

where y_{ep} is the output of product *p* by economy *e* in a given year.

Relative comparative advantage for being a transit point (RCAT-e)

The relative comparative advantage for being a transit point in global trade (RCAT-e) is the second filter used to determine the actual role of a provenance economy. This indicator represents the degree to which a given economy specialises in re-exporting pharmaceuticals, e.g. through development of advanced logistical infrastructure, or by its convenient geographical location. Consequently, it is assumed that such factors that facilitate transiting of genuine pharmaceutical products and medicines will also facilitate transit of fake pharmaceuticals.

The RCAT-e indicator is calculated by comparing relative volumes of re-export of pharmaceuticals to the shares calculated for other exporting economies. This is done based on re-export data that come from the UN Comtrade database (UN Trade Statistics Division, 2019).

Formally, the revealed comparative advantage in transit for an economy *e* within the pharmaceutical sector (RCAP-e) measures whether this economy re-exports pharmaceuticals as a share of its total manufacturing re-exports than the "average" country:

$$RCAT_{ep} = \frac{x_{ep} \Big/ \sum_p x_{ep}}{\sum_e x_{ep} \Big/ \sum_e \sum_p x_{ep}}$$

where x_{ep} is re-exports of product *p* by economy *e* in a given year.

Application of both filters

A complete list of RCAP-e and RCAT-e indices by economy can be found in Annex B.

Once the statistical filters (RCAP-e and RCAT-e indicators) are constructed, they are applied to distinguish the producing economies from the key potential transit points. Both filters are applied for every economy on the top provenance list for counterfeit pharmaceuticals, i.e. economies with a high GTRIC-e score (see Table 4.1 and Table 4.2).

The rationale for using the filters is as follows: if an economy is *not* a significant producer of fake pharmaceuticals (i.e. its RCAP is low) and/or is a large re-exporter of this good in legitimate trade of pharmaceuticals (i.e its RCAT is high), then it is likely to be a transit point.

On the other hand, if this top listed provenance economy of counterfeit pharmaceutical products and medicines is a significant producer (i.e. has a high RCAP score) or is a small re-exporter (i.e. has a low RCAT score), it is likely to be a producer of fake pharmaceuticals.

This exercise results in a list of producers and a list of transit points. Together with the information on the place of seizure, this will allow the development of maps of trade in fake pharmaceuticals, showing key producers, main transit point and main destination points.

Notes

[1] This is different to the economy's share of total imports of sensitive goods used to calculate GTRIC-p.

References

OECD/EUIPO (2016), *Trade in Counterfeit and Pirated Goods: Mapping the Economic Impact*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264252653-en.