

The study of doping: how to produce intelligence from Internet forums.

Thomas Pineau^{*a}, Adrien Schopfer^{*a}, Lionel Grossrieder^a, Julian Broséus^a, Pierre Esseiva^a,
Quentin Rossy^a

^aSchool of Criminal Justice
University of Lausanne
1015 Lausanne
Switzerland

*Corresponding authors: thomas.pineau@unil.ch & adrien.schopfer@unil.ch

Abstract

Despite the predominant role played by Internet in the distribution of doping substances, little is currently known about the online offer of doping products. Therefore, the study focuses on the detection of doping substances and suppliers discussed in Internet forums. It aims at having a comprehensive understanding of products and sellers to lead an operational monitoring of the online doping market.

Thirteen community forums on the Internet were investigated and one million topics were extracted with source code scrappers. Then, a semantic analysis was conducted with a semi-automatic process to classify the relevant words according to doping matters. Additionally, the ranking of doping products, active substances and suppliers in regards to the number of contributors to the forums were established and analysed over time. Finally, promotion methods of suppliers were evaluated.

The results show that anabolic androgenic steroids, used to enhance body image and performance, are the most discussed type of products. A temporal analysis illustrates the stability of the most popular products as well as the emergence of new products such as peptides (e.g. CJC-1295). 327 suppliers were detected, mostly with dedicated websites or direct sales by e-mail as selling methods.

Globally, the implemented methodology shows its ability to detect products and suppliers as well as to follow their temporal trends. The intelligence will serve the definition of online monitoring strategies (e.g. the selection of appropriate keywords). Additionally, it also allows the adjustment of customs inspection strategies and anti-doping analysis by monitoring the popular and emerging substances.

Keywords: illicit market, doping, scraping, intelligence-led screening, operational monitoring, forums

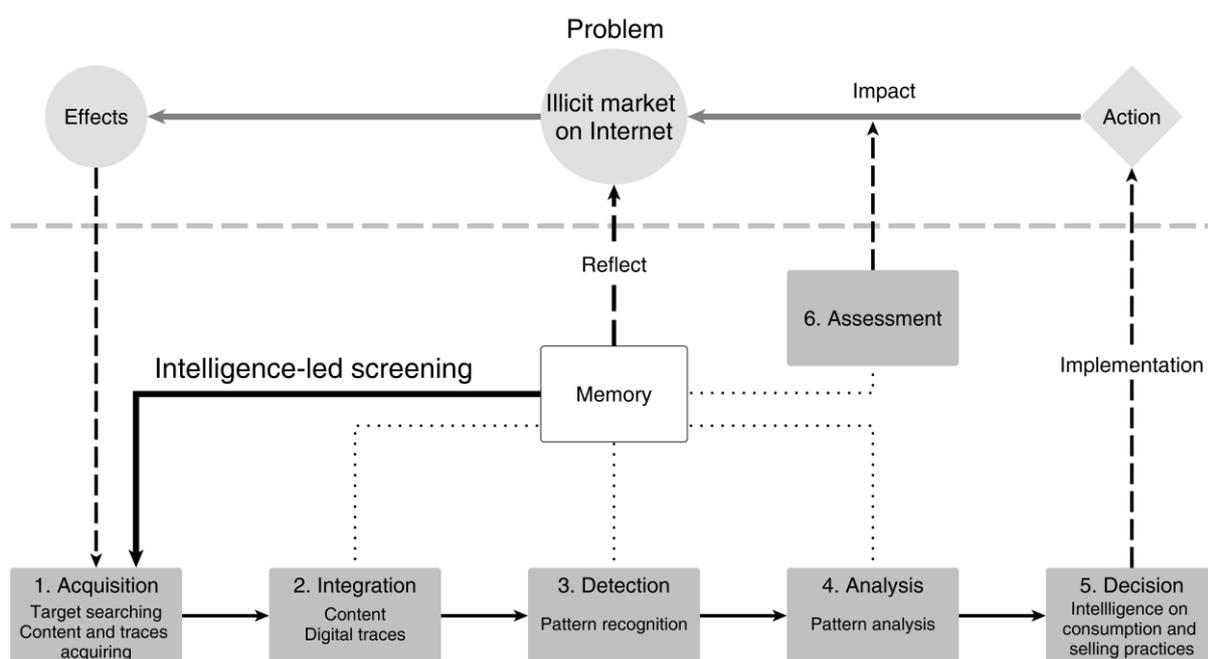
36 **Introduction**

37 The online market of doping products occupies a crucial place in the distribution and
38 accessibility of products [1,2]. The study of illicit online markets is a problem which can be
39 described in regards to usages of the Internet to support the market. Indeed, online services
40 may be used for several aims. They may be classified in three related processes: (1) the
41 promotion process which uses online services to advertise and ease the accessibility to
42 marketplaces, (2) the selling process which takes places in multiple online environments, and
43 (3) the discussion process which aims at sharing information and reviews about products and
44 sellers. If the selling process is central to the study of illicit online markets, promotion and
45 discussion processes are dependent issues that are also critical to tackle the global
46 phenomenon. Indeed, the discovery of illicit online markets requires to set up an appropriate
47 research strategy. This article focuses on the use of information gathered during the
48 discussion process to guide the search for online marketplaces.

49 Access to Internet sale spaces can take different paths, whether on the world wide web
50 including indexed websites (surface web) and not indexed websites (deep web) or overlay
51 networks (darknet) such as TOR.

52 For instance, search engines or directories can index webpages. Forums, social networking
53 services, sharing websites or spam may also be used to broadcast hyperlinks to webpages
54 or directly promote products [3,4]. Thus, the research of these promotion means requires the
55 identification of third-party services and information-sharing websites as well as the definition
56 of relevant keywords to perform target researches. Indeed, doing queries on different data
57 sources, whether they are search engines or the content of a website (e.g. social networking
58 services) is based on the use of keywords related to the problem to detect. Therefore, a set
59 of appropriate keywords according to the targeting content should be chosen to maximise the
60 efficiency of the process. This raises the question of keywords selection. These keywords
61 can be defined based on those used by already known sale websites (e.g. the keywords tag
62 in the source code or the textual content from the webpage). This strategy is limited because
63 it assumes that a sample of websites is already known and it is solely focused on indexed
64 websites. Moreover, the keywords obtained are not necessarily suitable for other online
65 communication spaces and unknown marketplaces. Other sources of information should thus
66 be used to guide the selection of keywords. The chosen approach in this research stems
67 from the study of online information sharing spaces, in particular on community forums on
68 the surface web, excluding deep web and darknet forums. The assumption is that the most
69 discussed terms found on community forums are the most suited to search websites or sale
70 announcements. Deep web and darknet forums were not taking into account in this study.
71 Indeed, doping is not considered as a very priority issue by law enforcements and can be
72 freely discussed on the surface web unlike pedophilia for example.

73 The described approach is comparable to the process named “Intelligence-led crime scene
 74 processing” formalised by Ribaux et al. [5,6]. This process aims at guiding the collection of
 75 forensic traces on a crime scene in the light of the knowledge previously accumulated about
 76 criminal activities. In particular, this knowledge is produced from the intelligence process
 77 which aims to detect and follow the evolution of crime problems [7]. The process is
 78 decomposed according to the following steps: (1) data collection; (2) integration in a
 79 structured memory; (3) detection of specific problems (through pattern recognition); (4)
 80 analysis of the detected problems; (5) intelligence production to guide the decision-making;
 81 and (6) impact assessment of the actions taken. The Figure 1, adapted from Ribaux [7],
 82 express the steps of the intelligence process in the case of illicit markets on the Internet.
 83



84
 85 **Figure 1** – Operational monitoring process in the case of illicit markets on the Internet.
 86

87 The arrow "intelligence-led screening", on the Figure 1, describes the research process of
 88 selling spaces on the Internet. This process is guided by the knowledge acquired during the
 89 data analysis of the crime problem. In this case, the analysis of community forums is used as
 90 a source of knowledge to identify and select the keywords needed to detect the selling
 91 spaces of doping products.

92 Forums analysis involves setting up a process that can be described by the previous
 93 approach. Thus, these discussion spaces are integrated in the global monitoring process of
 94 illicit markets on the Internet. The acquisition phase (1) includes a research of forums
 95 containing discussions about doping products, followed by an acquisition of their content.
 96 The acquired raw data includes general information about posts, members as well as the

97 promotional content (i.e. advertising, hyperlinks). The integration phase (2) aims at
98 structuring the relevant information by a semantic analysis of the topics to extract information
99 on products, active substances and suppliers. The detection phase (3) is mainly on the
100 identification of the most discussed terms and the identification of suppliers and their contact
101 means. The analysis step (4) focuses on the temporal analysis of detected products and
102 sellers in order to evaluate their trends and follow their evolution. Additionally, activity
103 patterns linked to selling contact means such as dedicated websites or e-mail are analysed.
104 Finally, (5) the intelligence inferred on selling and consumption practices aims at guiding
105 decision-making. In particular, to define relevant keywords to search the Internet for illicit
106 online marketplaces. Moreover, it allows the adjustment of custom inspections strategies and
107 anti-doping analysis by monitoring the popular and emerging substances.
108

109 **Previous studies**

110 A few studies have already examined the market of doping products on the Internet. Krug et
111 al. [8] have described the products that can be found on the market in Germany. The
112 products were seized by customs between 2010 and 2013. Steroids were the major kind of
113 seizures with 87.5% of the total number. In this category, esters of testosterone,
114 mehtandienone, boldenone and trenbolone were the main substances. Donati [1] pointed
115 that the world trafficking of doping products, in particular anabolic agents, is sharply
116 increasing through the Internet. In 2000, a first study was conducted to highlight the easiness
117 to buy doping products online [9]. However, the search strategy wasn't clearly described.
118 Additionally, Cordaro et al. [10], who did not provide information on their strategy of keywords
119 selection to detect websites, have investigated 30 marketplaces selling anabolic agents to
120 evaluate their availability on the Internet. They concluded that an average of 52 products
121 were available with esters of nandrolone, methandrostenolone and of testosterone in the
122 highest proportion. Clement et al. [11] studied the diffusion of nonprescription steroids on the
123 Internet. They selected 5 keywords related to the brand names of steroid products. They
124 classified 100 hyperlinks found with Google in 2006 to show the proportion of pro-use
125 websites according to keywords. Like Cordaro et al. [10] the strategy to select the keywords
126 is not described. Another study done by Brennan et al. [12] used Google to search selling
127 websites using generic keywords (like "steroids", "bodybuilding" and "growth hormone").
128 Unfortunately, there is a lack of details in the provided results (no classification of webpages,
129 no information about product available and no evaluation of keywords relevance). Moreover,
130 as the two previous studies, the authors did not mention the reasons for the keywords
131 choice.

132 Kraska et al. [13] did a sociological study of dealers in black markets. They analysed data
133 collected from several regional dealers who operated in fitness gyms. They described an

134 interesting mechanism where a dealer began first to buy steroid products from another
135 dealer then ordered on the Internet for his personal use to finally homebrewed, with raw
136 materials, his proper products and sell them. By spending 1,500\$ in raw material, the street
137 value of his products was around 50,000\$. This study suggests that dealing doping products
138 is highly profitable with a low risk compared to traditional illicit drugs such as heroin and
139 cocaine.

140 Moreover, according to Kraska et al. [13], Internet forums are one of the best sources to find
141 most of the information to produce or buy doping products, especially those used for
142 enhanced body image. These kind of forums called "Gray Web forum" by Wang et al. [14]
143 are used to encourage some deviant behaviour. As demonstrated by Seale et al. [15] and
144 Barrat [16], Internet forums are a valuable qualitative source of information. They are a "safe"
145 space where individuals can share and find information namelessly, in particular about their
146 own consumption of products. Therefore, researchers can find useful information related to a
147 sensitive topic such as illegal or immoral behaviour. Additionally, forums contain information,
148 which decrease the social desirability bias that often append with surveys. Forums provide
149 non-costly, unalterable data but also a broad temporality. On the other hand, this source
150 lacks of contextual information. Thus, the attributes of the population, such as gender, age or
151 situation, are difficult to evaluate. Community forums contain rich information that can be
152 collected for analytical purpose. The obvious limitation is how to process this large amount of
153 data. Manually, it would be extremely long to extract and process all the information.
154 Therefore, as proposed by Zhou et al. [17], an automatic tool must be built to manage this
155 task.

156

157 **Research objectives**

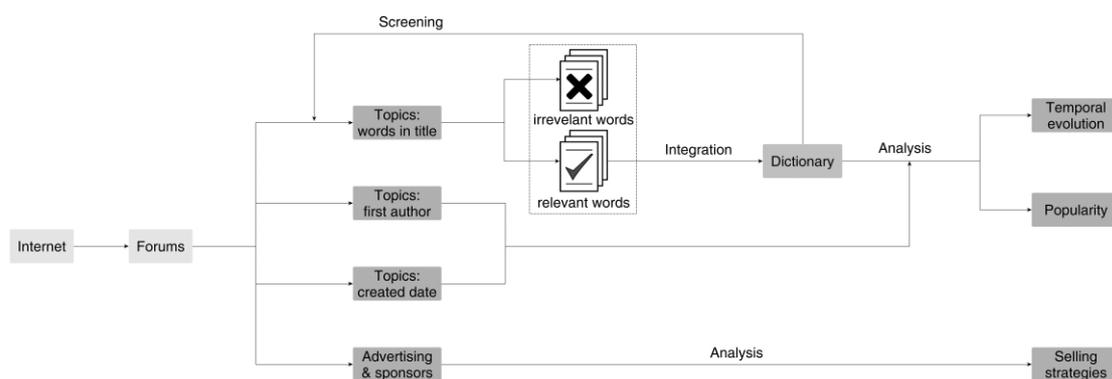
158 This research aims at analysing the discussions on online forums to identify doping products
159 and online sellers, as well as to evaluate their popularity and their temporal trends regarding
160 the number of users discussing them. In particular, the approach aims at detecting the most
161 discussed products and the emerging ones, which can be used as relevant keywords to
162 search the Internet for doping marketplaces.

163 Additionally, assessing the popularity of doping products could bring some insights about
164 their consumption, based on the hypothesis that the most discussed products are the most
165 consumed. Evaluating the popularity of doping products discussed online may also be useful
166 to settle or guide public health policy and prevention strategies.

167 Finally, the study also aims at developing a general methodology to extract and process
168 large amount of data available on forums. It is assumed that the approach may be reused for
169 the investigation of other problems and on other kinds of forums and topics.

170 **Methodology**

171 The global process is summarized in the Figure 2, and each step of the process is described
172 in the following sections. The first task is to detect online forums that contains discussions
173 about the relevant topic (in our case doping products). This may be done with a search
174 engine and by manually browsing the forums to identify their content. A general description
175 of their characteristics can then be performed (e.g. identifying their topics, number of users,
176 advertising and sponsors, etc.). Once forums are identified, it is then necessary to define
177 what information should be searched and collected. There are mainly two options: (1) to
178 browse all messages of the forum or (2) to browse the listings of topics for each subsection
179 of the forum. This choice should be guided by the research objectives. In our case, it was
180 assumed that listings of topics should be sufficient to identify products and sellers and to
181 evaluate their popularity. Indeed, one may assume that the first author of a topic would in
182 most cases indicate product and/or seller identification information in the title of his message.
183 Then a collection process has to be settled to browse all pages and crawl their content. The
184 downloaded listings are then parsed to integrate the information in a structured database
185 containing each relevant variable of listings such as titles, first author pseudonyms, dates,
186 etc. A semantic analysis of each word present in the title of the messages is then performed.
187 The process is structured in two steps: (1) an automatic filtering based on a stop-words list to
188 exclude words such as determinants, very common verbs or lexical words, numbers, etc. (2)
189 a systematic manual analysis of the reaming words to identify and classify them in categories
190 such as products, active substances, sellers, etc. Finally, a relevant indicator of popularity
191 should be chosen among the available indicators (i.e. the number of first authors, of topics, of
192 views or replies). This particular aspect is discussed in detail below.



193 **Figure 2 – Global methodology to analyse community forums.**
194

195 *Detection and selection of the forums*

196 To define our sample of Internet forums, a user-oriented approach was used. Keywords were
197 submitted in three search engines; google.com, bing.com and yahoo.com. The keywords
198 used were *steroid forum*, *growth hormone forum*, *peptide forum*, *fitness forum*, *aas forum*,
199 *doping forum*, *anabolic forum* and their translation in French. The keywords are generic
200 terms. They are only used to construct a sample of forums for this study. As stated in the
201 introduction, the process to discover new forums can be improved using the knowledge
202 acquired during the study.

203 The first fifty outputs were manually checked for each search engine. Only French and
204 English forums were chosen. This simple approach was used based on the hypothesis that it
205 may well correspond to the typical maximal research process undertaken by a
206 French/English-speaking person interested to find forums about steroids. The research was
207 performed in February 2015, and 27 forums were detected. The forums were classified
208 according to the presence of selling or doping products sections. Eight forums were excluded
209 because no sections on doping products were available. Indeed, the excluded forums are
210 mainly about dietary supplements, medical advice or training discussions. Additionally, 2
211 forums (muscegurus.com and forums.t-nation.com) were excluded due to their architecture
212 which was very difficult to monitor automatically. The remaining forums were filtered
213 according to the number of posts (arbitrary threshold set to more than 1000) and the number
214 of members (more than 500). At the end, 13 forums were kept for further analysis.

215 *Description of the forums*

216 The number of members and topics, the number of first-topic authors, the presence of
 217 specific seller sections or advertising banners in the forum and the existence of a dedicated
 218 sale website were extracted and reported manually (see Table 1).

219

Forums	Lang.	Number of first authors	Number of members	Number of topics	Number of adverts	Date of the first topic extracted (mm.yyyy)
steroid.com	EN	59028	308919	324985	8	08.2001
elitefitness.com	EN	39480	435373	276079	3	03.2000
steroidology.com	EN	29149	312894	128740	5	01.2003
ironmagazineforums.com	EN	10776	113338	56406	44	01.2001
thinksteroids.com	EN	10220	43487	46022	6	12.2003
steroids.com	EN	8818	132980	34348	3	10.2002
steroids.com	EN	6568	-	25362	0	05.2011
evolutionary.org	EN	3791	100818	21765	0	10.2009
anabolicsteroidforums.com	EN	3103	15385	23397	44	06.2012
steroid-forums.com	EN	2486	11612	11836	48	01.2011
ugbodybuilding.com	EN	1125	5641	4874	3	09.2011
fr.thinksteroids.com	FR	1022	7069	2801	7	09.2012
all-steroids.com	EN	580	4895	1878	12	09.2011
Total	-	176146	1492411	958493	183	-

220 **Table 1** – Selected forums with the number of members and topics, the number of selling
 221 adverts and sale websites affiliated to the forum.

222

223 Forums are more often than not subdivided in subsections that may globally be classified as:

- 224 - information about products and/or sellers (brand name, advises, efficacy, prices,
 225 delivery, scam, etc.)
- 226 - information about consumption practices (cycle, post-cycle, side effects, etc.)
- 227 - sponsors sections (specific sellers, advertising, etc.)
- 228 - security sections (“anonym” e-mail, VPN and proxy, etc.)
- 229 - others sections (training, women, over 40 years old, etc.)

230 The sections dedicated to the products are more often than not subdivided by the type of
 231 products such as steroids, growth hormones, peptides and supplements.

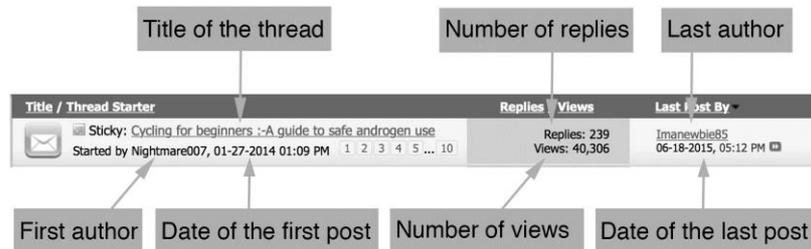
232

233 *Data collection and extraction process*

234 Each subsection contains a list of topics linking to a dedicated page containing the first
235 message sent and the replies. The first step in the process was to define a batch of URLs by
236 subsection for each forum. The main URL of each section was extracted. Then, a batch of
237 URLs specific to each section was automatically generated according to the number of
238 pages of the section. All the URLs were then stored in a text file used to feed an automatic
239 script performing the crawling to obtain the source code of webpages.

240 A dedicated extraction process was then developed to analyse all the sections containing the
241 lists of discussions (see Figure 3). The title of each thread, the first author, the date of the
242 first post, the last author, the date of the last post, the number of replies and the number of
243 views were extracted for each discussion in each forum section.

244



245

246

Figure 3 - Description of the extracted content.

247

248 A specific parser was developed to acquire the different targets embedded in the source
249 code. The identification of a stable structure and specific blocks for each forum led to the
250 development of only one extractor by forum. This step was performed using a scrapping tool:
251 import.io. This service enables to visually select the targeted structured content and find the
252 XPath of a specific content quickly. The dedicated extractors were run with a Python script
253 with all the batches of URLs. Finally, the extracted content was stored in csv files. At the end
254 of the acquisition step, 958'493 topics were extracted from the 13 Internet forums over 15
255 years.

256 The data were collected and processed in accordance with the Swiss Federal Act on Data
257 Protection (FADP). Indeed, the sources of data are publicly accessible online (without any
258 step of registration or any kind of protection) and do not contain sensitive personal data.
259 Additionally, the purpose of the study is not to identify specific peoples (i.e. virtual identities)
260 and only aims academic research purposes.

261 *Semantic analysis of the detected words*

262 Words contained in each topic title were separated from each other with a Python script
 263 using a regular expression and stored in a csv file for each forum. Then, the occurrence of
 264 each word was determined in all the topics previously extracted with no case sensitivity. The
 265 words with an occurrence smaller than five were excluded. To refine the list of words,
 266 irrelevant words were automatically eliminated using a stop-words dictionary (in French and
 267 English) provided by the Neuchâtel University¹ [18]. The remaining words were manually
 268 examined to determine if they were related to a product or a supplier as part of doping
 269 matters. The words identified as relevant were stored in a csv file referred as the "reference
 270 file". During this process, all words were described with the following metadata: name, class,
 271 category, active substance, group and contact information (see Tables 2 and 3).

272

Metadata	Description
Name	Full name related to the detected word (e.g. words: "sust" name: "sustanon")
Class	Word reference: product or supplier
Category	Type of product (e.g. steroid, growth hormone, EPO, etc.)
Group	Group of categories adapted from the WADA prohibited list (2015)
Active substance	Active substance present in the product
Contact information	Contact information (e-mail, websites, private message) for suppliers

273 **Table 2** – Description of the reference file metadata describing the detected words in all titles
 274 of the topics.

275

Word	Name	Class	Category	Group	Active substance	Contact
sus	sustanon	product	steroid	Steroid	testosterone	-
sus250	sustanon	product	steroid	Steroid	testosterone	-
susta	sustanon	product	steroid	Steroid	testosterone	-
sustanon	sustanon	product	steroid	Steroid	testosterone	-
ru	RUI-Products	supplier	peptide	Peptide, GF and EPO	-	http://www.rui-products.com/
ru -products	RUI-Products	supplier	peptide	Peptide, GF and EPO	-	http://www.rui-products.com/

276 **Table 3** – Example of the reference file describing the detected words in all titles of the
 277 topics.

278 At the end of the process, 1757 words related to doping were identified. The semantic
 279 analysis of these words allows their classification as 1167 different names wherein 158

¹ <http://members.unine.ch/jacques.savoy/clef/index.html> (accessed 14.06.2016)

280 active substances, 320 product brand names and 327 supplier names appeared. For each
281 supplier, a search in the posts was conducted to detect their contact information (e-mail and
282 dedicated websites).

283

284 *Choice of the indicator of popularity*

285 Several indicators of popularity for detected words detected may be chosen: the number of
286 topic views and responses, the number of topics containing a specific word, and the number
287 of distinct authors for each word. A Pearson correlation was calculated between the
288 indicators to assess if they measure the same popularity. All were positively correlated and
289 for 12 forums the correlations were higher than 0.641. The lower correlations were for French
290 forums where the total number of topics and forum members are smaller. This quantitative
291 observation tends to support the hypothesis that all indicators may be used to analyse the
292 popularity and temporal trends. However, qualitative aspects should also be taken under
293 consideration. In regards to the number of views, no information is available about the
294 counting process carried out by the forum and it could be different for one to another.
295 Regarding the number of responses, some topics are very popular such as tutorial and sticky
296 topics. On the other hand, seller locked topics (only accessible by suppliers) have a very low
297 number of responses. The number of topics is highly correlated with the number of distinct
298 first authors for each forum (from 0.986 to 0.999). Finally, the number of distinct first-authors
299 avoids the overestimation of the popularity of some words since specific authors may flood
300 the forums. Based on these considerations, the number of distinct first-authors was thus
301 chosen as the best indicator of popularity for active substances, brandnames and suppliers.
302 Nevertheless, it should be noted here that the question still remains whether the number of
303 distinct first authors may also be a relevant indicator of popularity of sellers or not. Indeed,
304 the presence of suppliers in discussions among different forums may be a better indicator of
305 popularity (see results). Thus, this indicator could avoid some flood effects due to the specific
306 activity of sellers.

307 *Trend analysis*

308 The temporal trends were evaluated using the date of the first post and the occurrences were
309 calculated using the number of distinct first author. The trends were reconstructed for the
310 groups of active substances, the brand names of products and the supplier's names. Only
311 words counted more than 25 times were selected. The trend analysis was thus performed on
312 290 different terms, and 515 terms were discarded.

313 To classify the different patterns of popularity over time, an exploratory approach based on a
314 trajectory analysis was used. In a second phase, the trends were checked manually to
315 confirm their classification. Among the different techniques available, a group-based
316 approach applied in developmental social science has been chosen [19–22]. Indeed, it has
317 proven its applicability on criminological issues with the longitudinal analysis of crime rates
318 along street segments [23,24]. By analogy, it appears to be applicable to our specific context.
319 The group-based trajectory model proposed by Nagin allows the identification of clusters of
320 temporal trends sharing similar patterns of evolution in time. As the approach is parametric,
321 several options were tested to identify the best model. The type and the order of the
322 statistical model has been determined by testing all the available options; Zero Inflated
323 Poisson, Censored Normal, and Logit forms (for the type) and as linear, quadratic, and cubic
324 forms (for the order). The results show that a Censored Normal quadratic model suits best
325 our data. To determine the number of trajectory groups, all combinations of the numbers of
326 groups (starting from three) and polynomial order of each trajectory were tested. The
327 Bayesian Information Criterion (BIC) was used to estimate the optimal number of groups
328 (Equation 1). The number of groups is increased by one and the models are compared two-
329 by-two to see which one fits best. This procedure continues until the model became unstable.
330 The software STATA 13 and the special procedure “Proc Traj” were used to perform the
331 analysis [25].

332

333
$$BIC = \log L - 0.5k \times \log N$$

334 **Equation 1:** Bayesian Information Criterion (BIC). L is the value of the model's maximized
335 likelihood estimates, N is the sample size, and k is the number of parameters.

336

337

338 **Results and discussion**

339 *Popularity of categories of products*

340 Among all first authors, 43% used at least once a word related to a doping product in their
341 topic titles. This suggests that the selected forums are highly specialized in doping matters.
342 The distribution of the different categories is quite similar in each forum. In other words, all
343 product categories are discussed similarly in each forum (see Table 4). There is no
344 specificity in spite of the different creation date of the forums and number of members. Thus,

345 communities seem to be interested in the same general subjects. This observation is
 346 supported by the comparison of the structure of the forums' sections.
 347 Steroid is the main discussed category. More than 30% of first topics authors wrote at least
 348 once about this category. This interest may be due to the long history of the usage of
 349 steroids, the broad products offer and a deeper knowledge (e.g. effectiveness, side effects,
 350 etc.). Peptide-growth factor-EPO, hormone-metabolic modulator and other anabolic agent
 351 categories are respectively mentioned by 9, 8 and 5% of topic authors. These four categories
 352 include all the doping products known to enhance performance and image (PIED).
 353

Type of substances	steroid.com	elitefitness.com	steroidology.com	ironmagazineforums.com	thinksteroids.com	isteroids.com	eroids.com	evolutionary.org	anabolicsteroidforums.com	steroid-forums.com	ugbodybuilding.com	fr.thinksteroids.com	all-steroids.com	All forums
Steroid	35.0	33.9	32.2	24.6	28.3	30.6	25.8	26.0	29.3	24.0	27.6	34.0	36.5	32.2
	8	1	8	4	9	2	4	1	9	1	4	5	5	6
Peptide, growth factor and EPO	10.0	8.44	8.61	7.93	9.54	8.26	11.1	7.39	10.1	12.5	11.2	8.71	8.45	9.21
	2						3		2	1	0			
Hormone and metabolic modulator	10.2	9.22	7.25	6.44	6.16	6.61	6.26	8.36	6.35	5.55	7.11	3.72	3.97	8.47
	6													
Other anabolic agent	5.98	4.87	3.26	2.94	3.00	4.37	1.39	4.17	2.93	2.61	2.22	5.58	4.48	4.50
Stimulant	1.42	2.11	1.02	2.00	1.14	0.57	0.14	2.19	1.45	0.93	0.27	1.57	1.55	1.44
Beta-2-agonist	0.35	0.39	0.27	0.17	0.38	0.59	0.05	0.69	0.23	0.16	0.36	0.59	1.03	0.34
Glucocorticoid	0.18	0.19	0.12	0.20	0.21	0.10	0.15	0.11	0.16	0.08	0.09	0.00	0.17	0.17
Diuretic and masking agent	0.12	0.29	0.12	0.01	0.18	0.10	0.06	0.08	0.10	0.00	0.00	0.10	0.00	0.15
Narcotic	0.08	0.36	0.02	0.04	0.03	0.05	0.15	0.13	0.13	0.04	0.09	0.00	0.00	0.13
Cannabinoid	0.01	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
All categories	46	46	41	41	41	40	38	41	39	37	39	45	47	43
Number of first authors	59'0	39'4	29'1	10'7	10'2	8'81	6'56	3'79	3'10	2'48	1'12	1'02	580	176'
	28	80	49	76	20	8	8	1	3	6	5	2		146

354 **Table 4** – Percentages of the number of distinct first authors talking about each category of
 355 products. Categories are adapted from WADA prohibited list (2015).

356 *Popularity of active substances*

357 The top 10 active substances of each forum were similar. Therefore, the results are
 358 presented and discussed for all forums (see table 5).

359

Steroids [%]	Peptides, Growth factor and EPO [%]		Hormones and metabolic modulators [%]		Others anabolic agents [%]		Stimulants [%]		
n = 176146	n = 176146		n = 176146		n = 176146		n = 176146		
testosterone	15.68	growth hormone	4.90	tamoxifen	2.67	clenbuterol	4.18	ephedrine	1.13
methandrostenolone	5.28	gonadotropin	3.43	anastrozole	2.32	enobosarm	0.28	caffeine	0.52
oxandrolone	4.98	igf	1.62	clomifene	2.18	ibutamoren	0.07	synephrine	0.12
nandrolone	4.81	ghrp	0.43	insulin	1.20	lgd	0.02	phentermine	0.07
trenbolone	4.81	melanotan	0.19	letrozole	1.18			methylphenidate	0.03
stanozolol	4.30	CJC	0.18	exemestane	0.66			selegiline	0.02
boldenone	2.88	thymosin	0.14	androstenedione	0.25			methylhexanamine	0.02
oxymetholone	1.63	EPO	0.11	cardarine	0.13			kratom	0.02
metenolone	1.56	mgf	0.10	arimistane	0.12			amphetamine	0.01
drostanolone	0.99	ghrh	0.05	toemifene	0.11			modafinil	0.01
dehydrochloromethyl-testosterone	0.96	pgf	0.04	raloxifene	0.11			sibutramine	0.01
mesterolone	0.60	triptorelin	0.03	formestane	0.06			cocaine	0.01
methasterone	0.73	gvs-111	0.01					dopamine	0.01
DHEA	0.54								
methyl-1-testosterone	0.35								

360

Beta-2-agonists		Glucocorticoids		Diuretics and masking agents		Narcotics	
n = 176146		n = 176146		n = 176146		n = 176146	
salbutamol	0.34	corticosteroid	0.13	spironolactone	0.08	ghb	0.12
		prednisone	0.03	furosemide	0.06	heroin	0.01
				glycerol	0.02		

361 **Table 5 – Percentages of first authors talking about each substance by active substances,**
 362 **for each class of doping product.**

363

364 Testosterone is the most represented substance among steroids. Only a few authors
 365 described precisely the composition of a product, therefore, all esters were merged for one
 366 substance. Testosterone is the endogenous anabolic steroid naturally present in the body.
 367 Moreover, the main goal of taking an anabolic agent is to produce more testosterone or
 368 mimic its body effect [26]. This may probably explain why this steroid is the most visible
 369 substance in forums. It also surpassed all the other categories of active substances in terms
 370 of the number of topics opened by authors. The other substances in the steroids class are

371 well known synthetic anabolic steroids. They have been involved in a lot of doping cases and
372 known to be effective [27]. Therefore, this is not surprising that they are popular in online
373 discussions.

374 Concerning the growth factors, peptides and EPO category, growth hormone is the most
375 represented substance. In forums, this substance is discussed as an alternative of steroids to
376 enhance muscle mass [28]. It stimulates many metabolites process such as secretion of
377 Insulin-like Growth Factor-I (IGF-1) and plays a key role in muscle protein synthesis. Growth
378 factors seem more popular than peptides and EPO, which may be explained by their
379 longstanding use compared to peptides. In the case of EPO, only 0.11% of first topic authors
380 spoke about it. It is likely due to the fact that EPO is used as blood doping agents in
381 endurance sports and do not have an effect on muscle mass. Therefore, this substance is
382 not worthwhile for fitness practitioners.

383 The substances in the other categories are much less popular. Except for ephedrine which is
384 used to lose weight, there is a low interest in illegal drugs such as amphetamine, cocaine,
385 heroin and cannabis. This may be explained by the fact that those substances are less
386 related to the field of bodybuilding or authors are less willing to discuss them publicly.

387 Globally, the most discussed substances are also the most used in bodybuilding according to
388 the review proposed by Kam and Yarrow [29]. Steroid, other anabolic agents and growth
389 factors are consumed to enhance muscle mass while hormone and metabolic modulators to
390 avoid side effects due to previous substances [2].

391

392 *Popularity of Brand names*

393 To analyse the popularity of brand names, all products with a brand name similar to an active
394 substance were removed from the ranking analysis (see Table 6). It has to be noted that a
395 brand name is not necessarily related to only one producer. For example, *Sustanon* is a
396 trademark owned by *Organon* but other producers such as *UKpharmalab* has taken the
397 same brand name.

Steroids n = 176146			Peptides, GF and EPO n = 176146			Hormones and metabolic modulators n = 176146		
Brandname	Active substance	[%]	Brandname	Active substance	[%]	Brandname	Active substance	[%]
anavar	oxandrolone	4.40	jintropin	hGH	0.17	nolvadex	tamoxifen	2.40
deca-durabolin	nandrolone	4.33	tb-500	peptide	0.17	clomid	clomiphene	2.32
winstrol	stanozolol	4.18	ipomorelin	GHRP	0.12	arimidex	anastrozol	1.25
danabol	methandrostenolone	4.13	somatropin	hGH	0.07	liquidex	anastrozol	0.67
sustanon	testosterone	3.50	serostim	hGH	0.10	aromasine	exemestane	0.66
equipoise	boldenone	2.62	pregnyl	gonadotropin	0.05	aromadex	exemestane	0.30
primobolan	metenolone	1.58	semorelin	GHRH	0.09	tamox	tamoxifen	0.13
anadrol	oxymetholone	1.41	genotropin	hGH	0.11	humalog	insulin	0.08
masteron	drostanolone	0.94	nutropin	hGH	0.03	formeron	formestane	0.03
proviron	mesterolone	0.94	saizen	hGH	0.04	teslac	testolactone	0.02
turinabol	dehydrochlormethyl-testosterone	0.25	hexarelin	GHRP	0.07	evista	raloxifen	0.01
anapolon	oxymetholone	0.08	kigtropin	hGH	0.06	liquinolva	tamoxifene	0.01
androtardyl	testosterone	0.01	igtropin	IGF	0.02	liquifem	Letrozole	0.01
			hygetropin	hGH	0.35			

Other anabolic agents n = 176146		
Brandname	Active substance	[%]
osta rx	clenbuterol	0.09
liquiclen	clenbuterol	0.04
superclen	clenbuterol	0.03
spiropent	clenbuterol	0.03
oxyflux	clenbuterol	0.02
ventipulmin	clenbuterol	0.02

398

399

400

Table 6 – Percentages of the number of first authors talking about each brand names by type of products.

401 In the steroid category, some brand names like Anavar, Sustanon, Winstrol, Deca-Durabolin
402 are originating from the pharmaceutical industry. Currently, pharmaceutical industries,
403 whether they are more or less legit depending of the country, also used those brand names
404 likely because of their quality reputation. For example, Deca-Durabolin is a brand name
405 owned by Aspen Global Incorporated (International registration: 693884A) which is used by
406 several suppliers, such as Organon, Teragon Labs, Meditech.

407 Brand names, such as Equipoise, come from the veterinary field (in higher doses than
408 pharmaceutical products) but they are used by humans. Globally, the brand names ranking
409 matches the previous ranking of active substances (see Table 5 and 6). Finally, forum
410 members refer more frequently to the brand name than directly to the active substance
411 name, in particular for stanozolol, methandrostenolone, nandrolone and oxandrolone.

412 Brand names of human growth hormone are the most represented among the group
413 "peptides, GF and EPO". Fewer differences are observed between the brand names of
414 human growth hormone. Therefore, consumers do not seem to have a particular preferred
415 product in this category.

416 Concerning brand names in the hormones and metabolic modulators category, the three
417 most popular are pharmaceuticals. This result supports the hypothesis that consumers like
418 (at least discuss) more pharmaceutical products than non-official products. This may be due
419 to their effectiveness and quality reputation.

420

421 *Popularity of Suppliers*

422 Unlike the brand names ranking or the active substances ranking, the suppliers' ranking is
423 specific to each forum. This is likely due to the presence of affiliate suppliers with specific
424 sections, advertising banner and/or selling part of the same website.

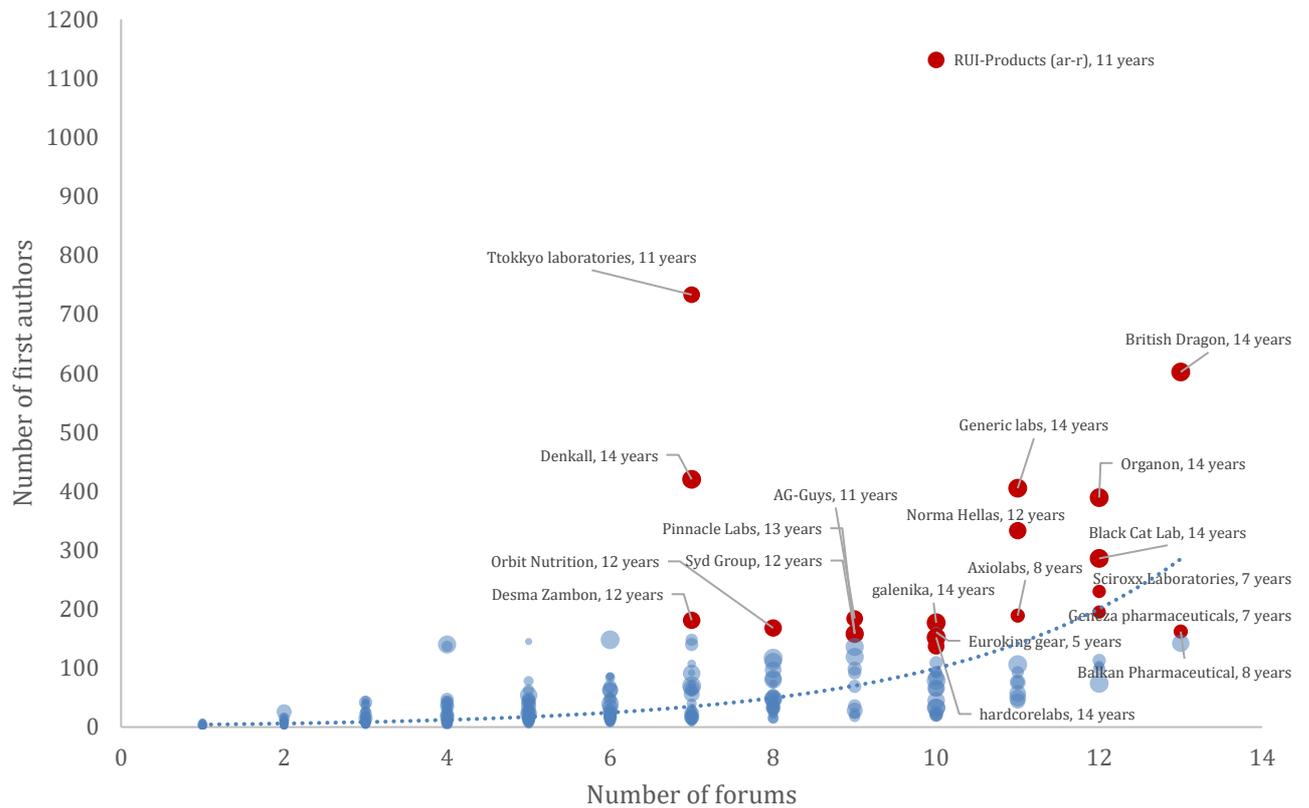
425 Table 7 shows the ranking of the 20 most discussed suppliers according to the number of
426 first topic authors. The numbers of first authors discussing suppliers increases exponentially
427 with the number of their apparition in distinct forums ($[\text{number of authors}] =$
428 $4.1743e^{0.3402[\text{number of forums}]}$, $R^2 = 0.936$, see Figure 4). Therefore, the presence of a supplier in
429 discussions among different forums seems to be a better indicator of popularity. As
430 previously stated, this indicator could avoid some flood effects due to the specific activity of
431 sellers. For example, RUI-products are discussed by more than 800 authors in only one
432 forum (accounting for 70% of all authors talking about these substances). These authors are
433 likely linked because they post exactly the same advertising thread. The observation time of
434 suppliers is defined as the time interval between the first thread and the last one where they
435 appear. The mean observation time is 5 years with a median of 4 years (see Figure 5). This
436 is a first indication about the life time of suppliers. However, it's not possible to verify if a

437 supplier brand name is used by several groups of suppliers. Additionally, consumers can
 438 discuss a supplier even after it stops its activity.
 439

Supplier	Authors	Forums	Observation time [years]
RUI-Products (AR-R)	1131	10	11
Ttokkyo Laboratories	733	7	11
British Dragon	602	13	14
Denkall	420	7	14
Generic Labs	405	11	14
Organon	389	12	14
Norma Hellas	333	11	12
Black Cat Lab	286	12	14
Sciroxx Laboratories	230	12	7
Geneza Pharmaceuticals	195	12	7
Axiolabs	189	11	8
AG-Guys	184	9	11
Desma Zambon	181	7	12
Galenika	177	10	13
Orbit Nutrition	168	8	12
Balkan Pharmaceutical	162	13	8
Euroking gear	161	10	5
Syd Group	159	9	12
Pinnacle Labs	158	9	13
Hardcorelabs	152	10	14

440

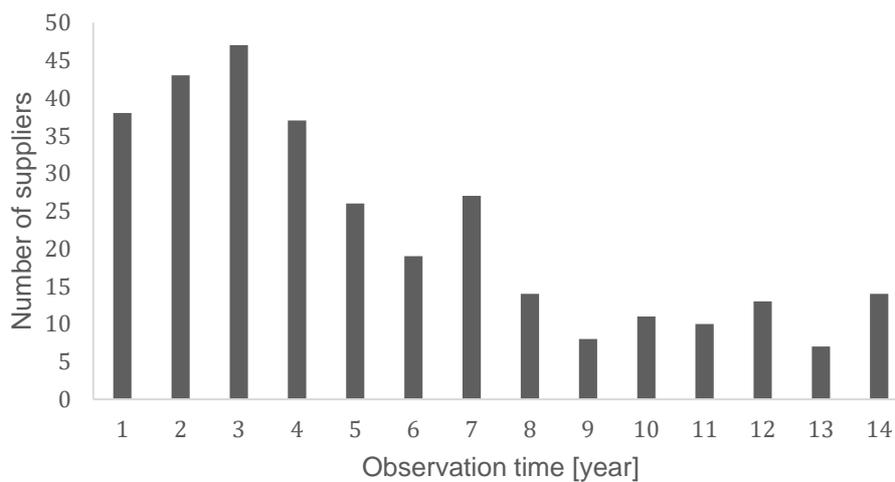
441 **Table 7–** Ranking of the top 20 suppliers according to the number of first authors, their
 442 presence on forums and the observation time.



443
444
445
446
447

Figure 4 – Popularity of the 327 suppliers according to their presence on forums and the number of first authors who discussed them. In red: the top 20 suppliers according to the numbers of first authors (see Table 7). The circle size is proportional to the observation time in years (see Table 7).

448
449



450
451
452
453

Figure 5 – Observation time of suppliers (n=327)

454 *Means of contact used by sellers*

455 Besides banners, three kinds of advertising strategies were detected in topic titles. The main
 456 strategy is to promote a selling website (n=260 suppliers). The second one is to give an e-
 457 mail address to contact the supplier (n=47 suppliers). Among the detected e-mails, the
 458 suppliers commonly used a provider selling a secured and anonymous service (Table 8).
 459 Finally, the last strategy is to contact the supplier through the private message system of the
 460 forum. Measuring this third strategy is difficult because any user can be contacted by a
 461 private message. The contact information was not detected for 80 suppliers because they
 462 seem to have disappeared.

463

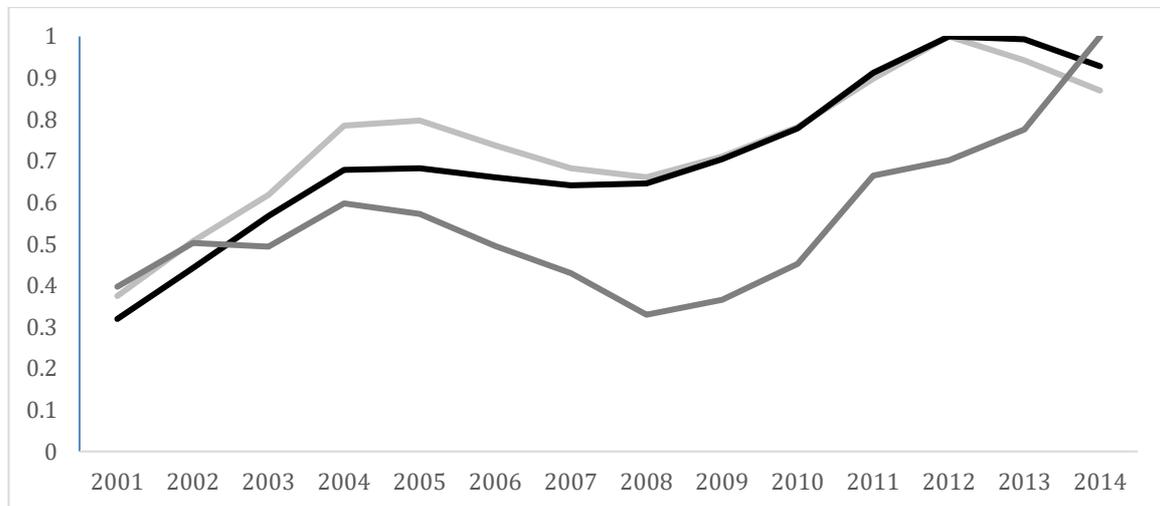
hostname	n
securenym.net	14
safe-mail-net	12
countermail.com	5
neomailbox.ch	2
others	14

464 **Table 8:** E-mail service detected (N=47)

465

466 **Trends analysis**

467 *Global trends*



468

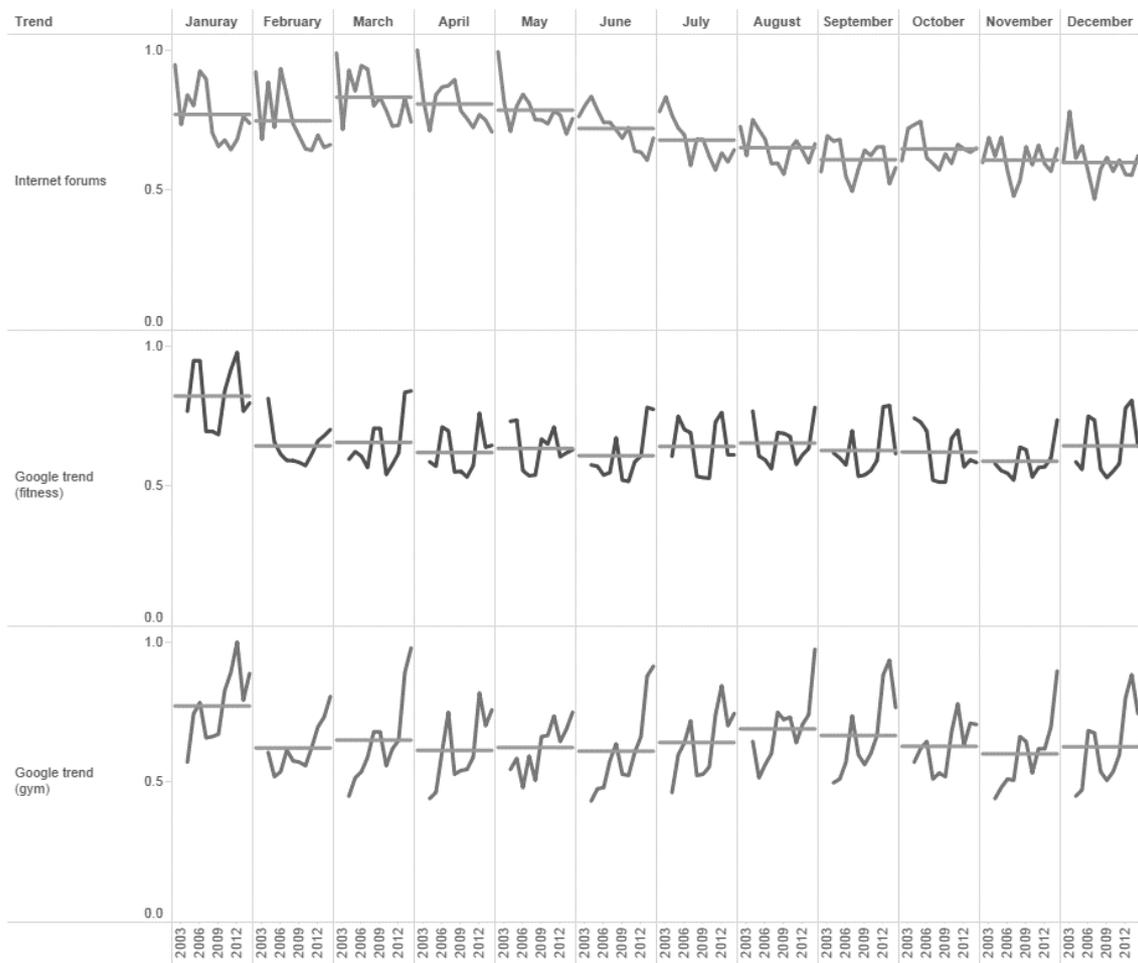
469 **Figure 6:** Temporal evolution of the number of first authors of topics. In black: the
 470 global/whole/full dataset (n = 176621), in middle grey: the supplier discussions (n = 13500),
 471 in light grey: the doping product discussions (n = 75559).

472

473 The Figure 6 presents the general posting trend on the 13 forums. The number of
 474 discussions about doping product topics is rising for the last 15 years according to the
 475 general trend. After the launch period of the three biggest forums (Table 1), the expansion of

476 active authors slowed down between 2005 and 2010. The same trend is observed for the
477 suppliers with a stronger decrease.

478 Several assumptions could explain this pattern. In 2005 and 2007, two operations were
479 conducted by the US Drug Enforcement Administration (DEA): “Operation Raw Deal” [30]
480 and “Operation Gear Grinder” [31]. As a result, 56 steroid laboratories and 8 companies were
481 identified and their products seized. These events could explain a diminution of the visibility
482 of suppliers. Indeed, some members included the suppliers could be more careful, to avoid
483 being targeted. On the other hand, the appearance and increasing importance of social
484 networking services (e.g. Facebook in 2004) could bring a displacement of the forums
485 members to a new trendy way of communication.
486



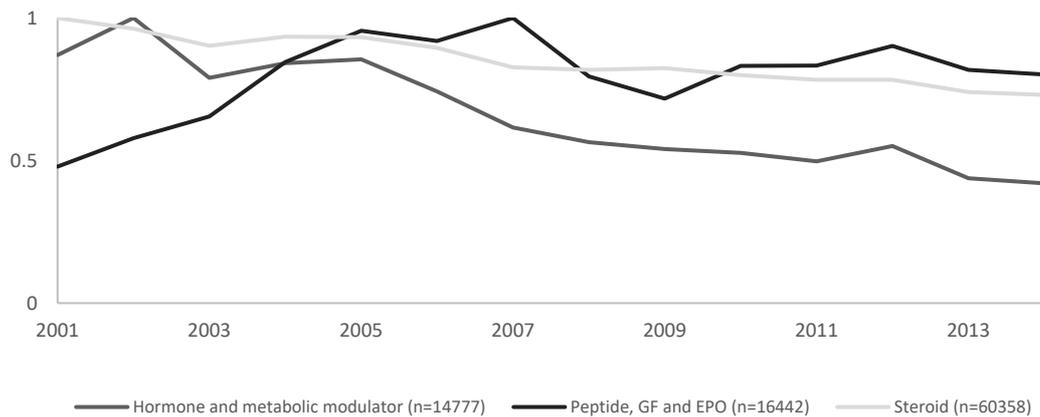
487
488 **Figure 7: Seasonality of posting for the 3 biggest forums.**
489

490 The seasonality is presented in Figure 7. The number of topics is bigger at the beginning of
491 the year (January to May). This seasonal trend may be due to some “good New Year’s

492 resolutions". This trend is confirmed by data collected through Google trends² that shows the
493 same cyclical trend for January (with 'gym' and 'fitness' as keywords). The seasonal trend
494 makes us assume that the main interest of the forum's members is the use of doping agents
495 in a fitness context. Indeed, Read [32] concluded that 75% of all gym memberships are taken
496 out in January.

497

498 *Categories of doping products*



499

500 **Figure 8:** General trends for the 3 most popular categories.

501

502 Among categories, the hormones and metabolic modulators category has a downward trend
503 from 2002 to 2014 (see Figure 8). The steroids category has a chronic trend (always high)
504 slightly decreasing. This can be due to a saturation effect. Indeed, the number of discussions
505 could decline because of the large number of topics already posted which may answer the
506 majority of the members' enquiries. Additionally, a displacement of the use of a specific type
507 of products could also partially explain these trends.

508 Discussions about peptides and growth factors increase since 2001 with a maximum in 2007.
509 Compared to other categories of substances, they are the new generation of doping agents
510 [33]. Members of forums seem to speak a lot about them and continuously open new topics
511 to answer their questions. In addition, these substances have a reputation, inside the forums
512 community, to be medically safer with none or fewer side effects than anabolic steroids.

² <https://www.google.ch/trends> (accessed 13.06.2016)

513 *Products, brand names and suppliers*

514 The trend analysis was performed on 290 different terms, 515 terms being discarded, due to
 515 their low occurrences ($n < 25$). The number of active substances, brand names and suppliers
 516 analysed can be found in table 9.

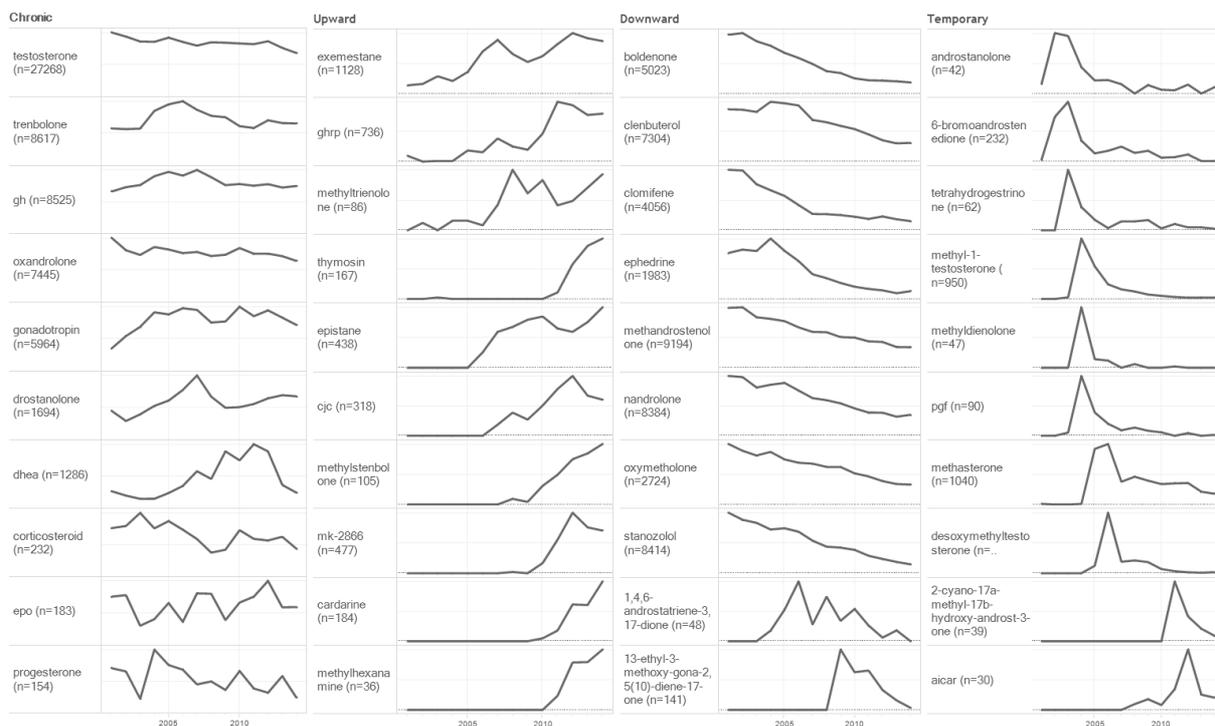
517 The group-based trajectory analysis classified trends in 4 groups: upward trends, chronic
 518 trends, downward trends and temporary increase. In a second phase, these groups were
 519 checked manually to confirm their classification. The automated method correctly classified
 520 them depending on the type of trend: chronic trends at 34%, upward trends at 75%;
 521 downward trends at 76% and temporary increases at 65%. Chronic trends are misclassified
 522 due to the overlaps with the temporary increase type. This automated method saved time
 523 and helped to define the main groups of the trends. However, results show that a manual
 524 verification is still necessary to ensure a reliable classification. The results are summarized in
 525 Table 9.

	Chronic	Upward	Downward	Temporary increase	<i>n</i>
Substances	14%	20%	32%	34%	97
Brandnames	8%	8%	48%	36%	132
Suppliers	2%	54%	18%	26%	61

527 **Table 9:** Percentage of the types of trends.

528

529 A sample of trends of active substances, brand names and suppliers in each of the four
 530 trends categories are respectively shown in Figures 9 to 11.



531 **Figure 9:** Sample of 40 trends for active substances ($n = 97$).

532

533

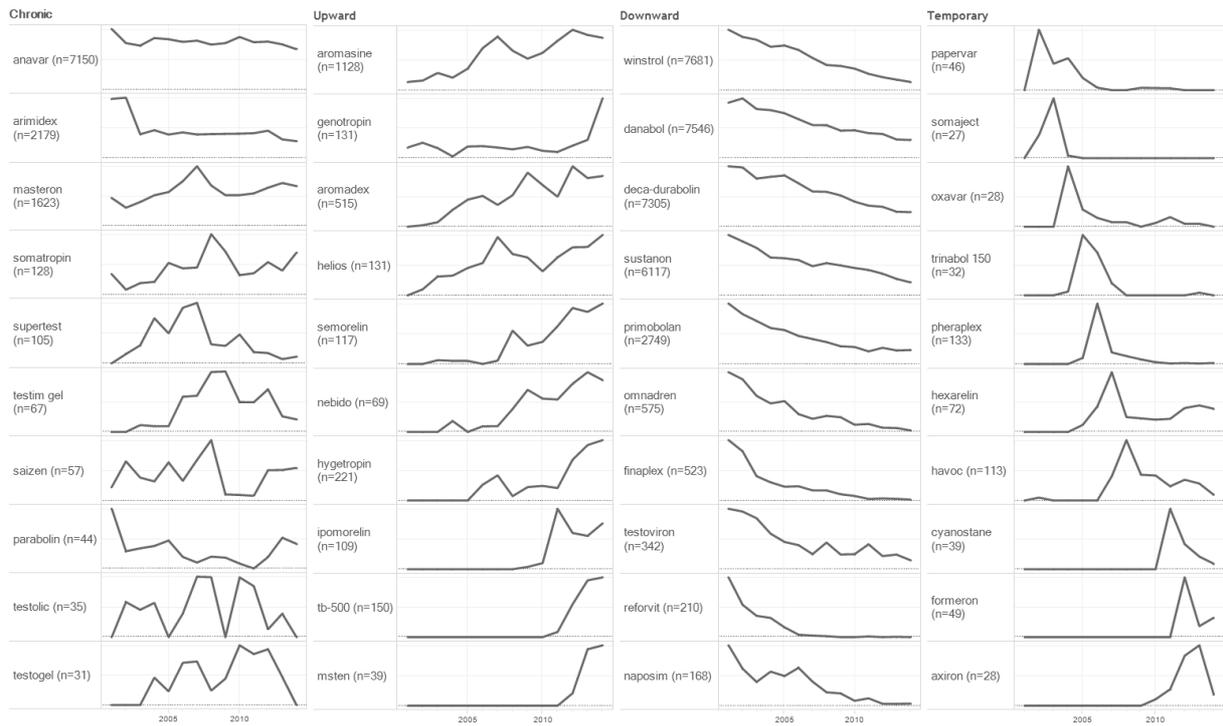


Figure 10: Sample of 40 trends for brand names (n = 132).

534

535

536

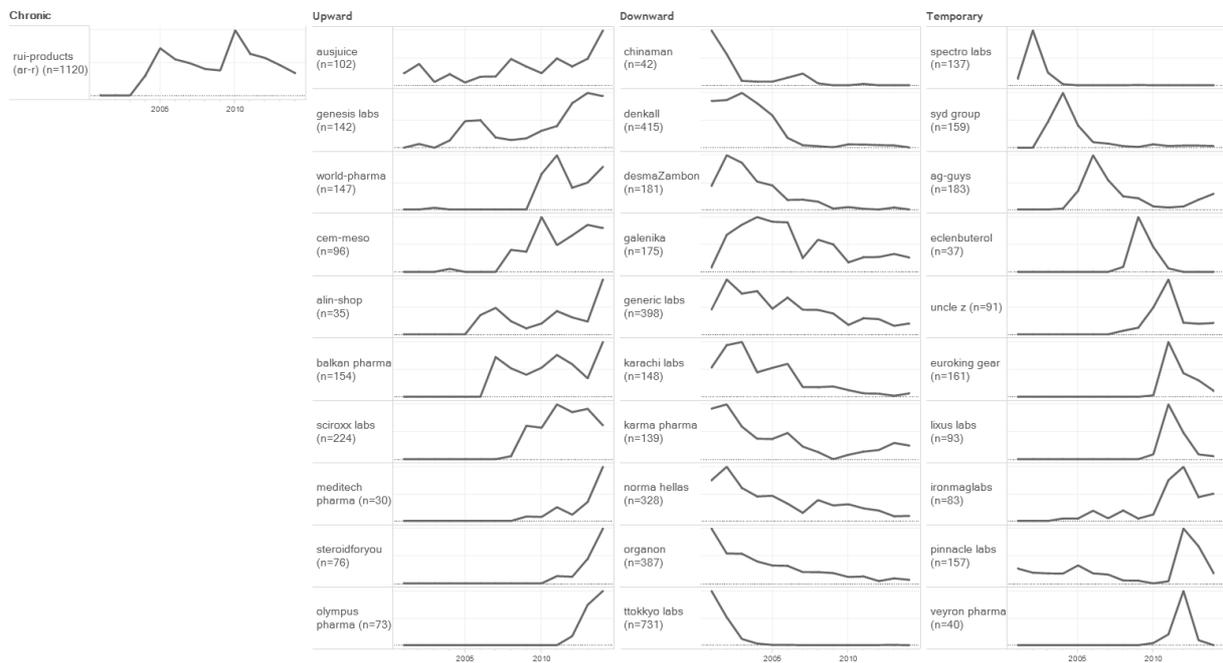


Figure 11: Sample of 31 trends for suppliers (n = 61).

537

538

539

540

541

542

543

Several assumptions can be drawn up about the trends. Upward trends could be due to the growing popularity of a product or a supplier as a result of customer satisfaction, new entry on the market or in the prohibited list edited by the WADA. For example, Ostarine (mk-2866) is a selective androgen receptor modulator (SARM) which was banned by WADA in 2008.

544 The Swiss laboratory for Doping Analyses reported the first case of abuse involving SARMS
545 in 2010 [34]. This product showed an upward trend since 2010 (year of its first detection in
546 community forums, see Figure 9). It is worth noting that only few brand names, substances
547 or suppliers showed an upward trend (63 on 290).

548 Chronic trends could be due to the quality and the reliability of products and suppliers
549 respectively. Additionally, the promotional strategies of suppliers could also impact the trend
550 as shown below. The majority of the chronic trends is stable since 2001. This stability
551 suggests that products or suppliers are well established in the market and therefore have a
552 strong reputation or an intense promotional strategy. For example, testosterone is the leader
553 anabolic steroid on the market and has a chronic trend since 2001. RUI-product also known
554 as AR-R is the only supplier showing a chronic trend since 2004. The promotion of this
555 supplier in the community forums was very extensive with many different usernames.

556 Downward trends could be due to customers' dissatisfaction, saturation effect, market
557 competition, less accessibility or diversity of supply. For example, the majority of the most
558 popular substances or brand names such as danabol, deca-durabolin, clenbuterol, clomifene
559 and pgf showed downward trends likely due to a saturation effect. As previously explained,
560 large number of topics are already in place answering members' enquiries.

561 Finally, temporary increases could be due to media effect, sudden disappearance such as a
562 laboratory shutdown or advertising campaign. For example, tetrahydrogestrinone (THG) is an
563 anabolic steroid with a maximal popularity in 2003. In this year, it was firstly detected when a
564 syringe containing the drug was given to the U.S. Anti-Doping Agency (USADA) [35].
565 Consequently, a number of athletes were tested positive for THG. The THG scandal and the
566 increase in popularity are concomitant. Another example is e clenbuterol which was a supplier
567 selling through the website "www.eclenbuterol.com". Its website was opened in 2008 and
568 shutdown the last semester of 2009 according to the forums posts and www.archive.org (last
569 access: May 2016). The temporary trend corresponded to the short observation time of this
570 supplier with a majority of topics concerning its disappearance.

571 Substances are mainly classified in downward or temporary increased categories. The
572 substances with an occurrence (n) higher than 2000 are chronic or slightly downward. Brand
573 names are much related to the substances. Therefore, they present the same pattern of
574 trends. However, suppliers are mostly classified in the upward category with a start after
575 2011 and the temporary increases last a maximum of three years. Additionally, only one
576 supplier was chronic. Therefore, suppliers seem to have a short lifetime probably due to the
577 low stability of the market.

578

579 **Conclusion and perspectives**

580 The study focuses on the detection of doping substances and suppliers discussed in Internet
581 forums to acquire a comprehensive understanding of the online market. 13 community
582 forums on the Internet were investigated and nearly one million topics were extracted with
583 source code scrappers.

584 The results show that the most discussed type of products is anabolic androgenic steroids
585 which is used to enhance body image and performance. The most popular products are
586 stable over time, and the emergence of new products such as peptides is observed. 327
587 suppliers were detected with mostly dedicated websites or direct sales by e-mail as selling
588 methods. Globally, the proposed methodology shows its ability to detect products and
589 suppliers and follow their temporal trends.

590 The generated intelligence will serve to guide the searching process, presented as
591 intelligence-led screening, of the online doping marketplaces based on the most popular
592 terms (i.e. keywords) used for both active substances and brand names. Additionally, it
593 allows the enhancement of customs inspection strategies and anti-doping analysis by
594 monitoring the most discussed and emerging substances. For example, a list of emerging
595 substances or the reference file (containing information such as the names of active
596 substances, brands or suppliers) may be provided to actors involved in anti-doping regulation
597 to improve their knowledge. Additionally, trends can be monitored in real time and accessible
598 through a community platform or a dashboard. These trends can be used to assess
599 operational decision or prevention on a specific substance, category or supplier.

600 As limitations, this study focused only on indexed forums on the surface web. Darknet and
601 deep web forums could be investigated to obtain a better overview. Moreover, discussions in
602 topics have not been taken into account. Thus, some words related to doping matter may be
603 missed. Compounds terms also were not detected, except for some names of laboratories
604 (e.g. British Dragon). Additionally, the popularity is only based on the number of first-topic
605 authors using the words. Therefore, positive or negative opinions were not taken into
606 account. Sentiment analysis could be performed in a second step to obtain a more precise
607 assessment about the popularity of products and sellers. Moreover, a study of other medium
608 of communication (such as social networking services) could bring additional and
609 complementary information about the popularity and sales of doping products.

610 As perspectives, trends comparison with customs and law enforcement seizures as well as
611 WADA-accredited laboratories results is currently ongoing to obtain a cross-validation of the
612 popularity of doping products. The comparison also aims at evaluating the hypothesis that
613 the most discussed products are the most consumed. Additionally, since May 2015, the most
614 popular terms identified in this study that are related to steroids are currently used in a

615 systematic monitoring process on search engines and social networking services to detect
616 websites selling doping products.
617 Finally, it is assumed that the described methodology can be widely applied to other types of
618 crime problems and security issues. For example, the monitoring of extremist groups in
619 forums, market analysis, and other research purposes in forum communities.

620 **Bibliography**

- 621 [1] A. Donati, World traffic in doping substances, 2007.
- 622 [2] F. Sjöqvist, M. Garle, A. Rane, Use of doping agents , particularly anabolic steroids ,
623 in sport and society, *Lancet*. 371 (2008) 1872–1882.
- 624 [3] B.J. Jansen, A. Spink, How are we searching the World Wide Web? A comparison of
625 nine search engine transaction logs, *Inf. Process. Manag.* 42 (2006) 248–263.
626 doi:10.1016/j.ipm.2004.10.007.
- 627 [4] R. Koopmans, A. Zimmermann, Visibility and communication networks on the internet:
628 The role of search engines and hyperlinks, in: *A Eur. Public Sph. How Much It Do We
629 Have How Much Do We Need?*, MZES, 2007: pp. 213–264.
- 630 [5] O. Ribaux, A. Baylon, C. Roux, O. Delémont, E. Lock, C. Zingg, P. Margot,
631 Intelligence-led crime scene processing. Part I: Forensic intelligence., *Forensic Sci.
632 Int.* 195 (2010) 10–6. doi:10.1016/j.forsciint.2009.10.027.
- 633 [6] O. Ribaux, A. Baylon, E. Lock, O. Delémont, C. Roux, C. Zingg, P. Margot,
634 Intelligence-led crime scene processing. Part II: Intelligence and crime scene
635 examination., *Forensic Sci. Int.* 199 (2010) 63–71. doi:10.1016/j.forsciint.2010.03.011.
- 636 [7] O. Ribaux, *Police scientifique - Le renseignement par la trace*, PPUR presses
637 polytechniques, 2014.
- 638 [8] O. Krug, A. Thomas, K. Walpurgis, T. Piper, G. Sigmund, W. Schänzer, T.
639 Laussmann, M. Thevis, Identification of black market products and potential doping
640 agents in Germany 2010-2013., *Eur. J. Clin. Pharmacol.* 70 (2014) 1303–11.
641 doi:10.1007/s00228-014-1743-5.
- 642 [9] V. Dumestre-Toulet, Doping via the Internet ? ...a mouse game!, *Ann. Toxicol. Anal.*
643 12 (2000) 19–25.
- 644 [10] F.G. Cordaro, S. Lombardo, M. Cosentino, Selling androgenic anabolic steroids by the
645 pound: identification and analysis of popular websites on the Internet, *Scand. J. Med.
646 Sci. Sport.* 21 (2011) 247–259. doi:10.1111/j.1600-0838.2010.01263.x.
- 647 [11] C.L. Clement, D.B. Marlowe, N.S. Patapis, D.S. Festinger, R.F. Forman,
648 Nonprescription steroids on the Internet., *Subst. Use Misuse.* 47 (2012) 329–41.
649 doi:10.3109/10826084.2011.630225.
- 650 [12] B.P. Brennan, G. Kanayama, H.G. Pope, Performance-enhancing drugs on the web: a
651 growing public-health issue., *Am. J. Addict.* 22 (2013) 158–61. doi:10.1111/j.1521-
652 0391.2013.00311.x.

- 653 [13] P.B. Kraska, C.R. Bussard, J.J. Brent, Trafficking in Bodily Perfection: Examining the
654 Late-Modern Steroid Marketplace and Its Criminalization, *Justice Q.* 27 (2010) 159–
655 185. doi:10.1080/07418820902814013.
- 656 [14] J.-H. Wang, T. Fu, H.-M. Lin, H. Chen, Exploring Gray Web Forums: Analysis and
657 Investigation of Forum-Based Communities in Taiwan, in: *Stud. Comput. Intell.* Vol.
658 135, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008: pp. 121–134.
659 doi:10.1007/978-3-540-69209-6_7.
- 660 [15] C. Seale, J. Charteris-Black, A. MacFarlane, A. McPherson, Interviews and internet
661 forums: a comparison of two sources of qualitative data., *Qual. Health Res.* 20 (2010)
662 595–606. doi:10.1177/1049732309354094.
- 663 [16] M.J. Barratt, Discussing illicit drugs in public internet forums : Visibility , stigma , and
664 pseudonymity, *C T.* (2011) 159–168. doi:10.1145/2103354.2103376.
- 665 [17] Y. Zhou, J. Qin, G. Lai, H. Chen, Collection of U . S . Extremist Online Forums : A Web
666 Mining Approach Department of Systems and Industrial Engineering , University of
667 Arizona Department of Management Information Systems , University of Arizona,
668 (2007) 1–10.
- 669 [18] Z. Yao, C. Ze-Wen, Research on the construction and filter method of stop-word list in
670 text preprocessing, *Proc. - 4th Int. Conf. Intell. Comput. Technol. Autom. ICICTA*
671 2011. 1 (2011) 217–221. doi:10.1109/ICICTA.2011.64.
- 672 [19] D. Nagin, K. Land, Age, criminal careers, and population heterogeneity: Specification
673 and estimation of a nonparametric, mixed Poisson model, *Criminology.* 31 (1993)
674 327–362. doi:10.1111/j.1745-9125.1993.tb01133.x.
- 675 [20] D.S. Nagin, D.P. Farrington, T.E. Moffitt, Life-Course Trajectories of Different Types of
676 Offenders, *Criminology.* 33 (1995) 111–139. doi:10.1111/j.1745-9125.1995.tb01173.x.
- 677 [21] D.S. Nagin, Analyzing developmental trajectories. A semi-parametric group based
678 approach, *Psychol. Methods.* 6 (1999) 18–34. doi:10.1037/1082-989X.4.2.139.
- 679 [22] D.S. Nagin, *Groupe-Based Modeling of Developpment*, Harvard University Press,
680 2005.
- 681 [23] D. Weisburd, S. Bushway, C. Lum, S.-M. Yang, Trajectories of Crime at Places: A
682 Longitudinal Study of Street Segments in the City Of Seattle, *Criminology.* 42 (2004)
683 283–321. doi:10.1111/j.1745-9125.2004.tb00521.x.
- 684 [24] D. Weisburd, E.R. Groff, S.-M. Yang, *The criminology of place*, Oxford university
685 press, 2012.
- 686 [25] B.L. Jones, D.S. Nagin, K. Roeder, *A SAS Procedure Based on Mixture Models for*

- 687 Estimating Developmental Trajectories, *Sociol. Methods Res.* 29 (2001) 374–393.
688 doi:10.1177/0049124101029003005.
- 689 [26] T.D. Fahey, *Anabolic-Androgenic Steroids: Mechanism of Action and Effects on*
690 *Performance*, *Encycl. Sport. Med. Sci.* (1998).
691 <http://www.sportsci.org/encyc/anabster/anabster.html> (accessed July 27, 2016).
- 692 [27] R. Kazlauskas, *Designer Steroids*, in: *Doping Sport.*, First Edit, Springer, 2010: pp.
693 155–186. doi:10.1007/978-3-540-79088-4.
- 694 [28] M. Saugy, N. Robinson, C. Saudan, N. Baume, L. Avois, P. Mangin, *Human growth*
695 *hormone doping in sport.*, *Br. J. Sports Med.* 40 (2006) 35–39.
696 doi:10.1136/bjism.2006.027573.
- 697 [29] P.C. Kam, M. Yarrow, *Anabolic steroid abuse: Physiological and anaesthetic*
698 *considerations*, *Anaesthesia.* 60 (2005) 685–692. doi:10.1111/j.1365-
699 2044.2005.04218.x.
- 700 [30] Department of Justice, *International Law Enforcement Operation Targets*
701 *Underground Manufacture of Anabolic Steroids*, (2007).
702 https://www.justice.gov/archive/opa/pr/2007/September/07_crm_753.html (accessed
703 July 27, 2016).
- 704 [31] Drug Enforcement Administration, *DEA Leads Largest Steroid Bust in History*, (2005).
705 <https://www.dea.gov/pubs/pressrel/pr121505.html> (accessed July 27, 2016).
- 706 [32] T. Read, D. Davies, *The Secret Life of Money: Everyday Economics Explained*, First
707 Edit, London, 2015.
- 708 [33] Australian Crime Commission, *Organised Crime and Drugs in Sport*, 2013.
- 709 [34] E. Grata, L. Perrenoud, M. Saugy, N. Baume, *SARM-S4 and metabolites detection in*
710 *sports drug testing: A case report*, *Forensic Sci. Int.* 213 (2011) 104–108.
711 doi:10.1016/j.forsciint.2011.07.014.
- 712 [35] L.D. Bowers, *Anti-dope testing in sport: The history and the science*, *FASEB J.* 26
713 (2012) 3933–3936. doi:10.1096/fj.12-1001ufm.
- 714