

# Paper computationele biologie: Metabolomics

Elias El Bouzidi  
Amber Van den bosch  
Yassine El Abdellati  
Philippe voet

May 2024

## 1 Inleiding

### 1.1 Wat is Metabolomics?

Metabolomics is een relatief nieuw onderzoeksdomein dat focust op het metaboloom, de gehele verzameling van metabolieten in een afgebakend biologisch systeem zoals een organisme. Een metaboliet is een kleine organische molecule die geproduceerd, gewijzigd wordt of gewoon deelneemt aan biochemische processen die plaatsvinden in een biologisch systeem om het in leven te houden. De definitie van ‘klein’ omvat meestal moleculen tot 1500 Da, waardoor aminozuren er nog wel bijgeteld worden, maar hele eiwitten niet. De metabolieten kunnen worden opgesplitst in endogene stoffen en exogene stoffen, afhankelijk of ze in of buiten het (menselijk) lichaam worden geproduceerd. Drugs zijn voorbeelden van exogene metabolieten.

### 1.2 Technieken in Metabolomics

Om aan metabolomics te kunnen doen, is het belangrijk om metabolieten uit samples te kunnen identificeren en hun concentratie te weten te komen. Identificatie kan bereikt worden door massaspectrometrie toe te passen. In de praktijk wordt er vaak gebruik gemaakt van LC-MS/MS, wat wil zeggen dat er voor de scheiding van de verschillende metabolieten Liquid Chromatography wordt toegepast en dat de eigenlijke massaspectrometrie in tandem gebeurt. Hierbij worden er twee spectra bekomen, één van de volledige geïoniseerde metabolieten en één van de gefragmenteerde stukken daarvan. Uit de massa over lading – wat in het geval van metabolieten vaak 1 is, waardoor de massa direct bekend is – verhouding die deze spectra weergeven, in combinatie met de intensiteit van de ‘peaks’, kan de massa bepaald worden en de onbekende metaboliet gekoppeld worden met de hoogst scorende match uit een soort van referentiebibliotheek, waarin spectra zitten van vooraf geïdentificeerde metabolieten. Deze matches worden a.d.h.v. een aantal criteria beoordeeld (bv. de cosine similarity, wat een

coëfficiënt is die meer zegt over de overlap tussen de twee spectra). Toch verloopt dit proces niet altijd vloeiend, waardoor er tot voor kort maar vijf tot tien procent van de spectra die uit stalen kwamen geïdentificeerd kon worden. Met de toepassing van AI op dit proces kon deze succesrate worden opgeschroefd.

### **1.3 Heranalyse van Oude Data**

Met de komst van deze vernieuwde spectrumanalyse technieken, kan het interessant zijn om data van oude experimenten opnieuw te gaan analyseren. Dit wordt mogelijk gemaakt door platformen zoals GNPS en ReDU, welke ervoor zorgen dat in tandem-massaspectrometrie data uit openbare experimenten alsook hun metadata worden opgeslagen zodat re-analyse kan gebeuren.

### **1.4 Projectdoelstellingen**

In dit project wordt er ook gebruik gemaakt van re-analyse van publieke data. Initieel werd er getracht verbanden te leggen tussen spectra van illegale drugs uit samples en wereldhavens. Er zouden vermoedelijk geografische verschillen optreden, welke we wilden onderzoeken en indien mogelijk een link maken met het waarom. Dit bleek echter niet mogelijk te zijn met de beschikbare data, er konden te weinig drugs aan havensteden gelinkt worden. Daarom werd de omschakeling naar een hieraan gerelateerde onderzoeksvraag: ‘In welke types van samples worden de meeste illegale drugs teruggevonden?’ Hiervoor wordt er gebruik gemaakt van bestanden afkomstig uit GNPS-analyses (acht verschillende searches), de metadata van deze experimenten opgeslagen in ReDU en meer informatie over de verschillende drugs afkomstig van DrugBank. Al deze data worden samengebracht en door een algoritme geanalyseerd op mogelijke links tussen de relevante metadata. Vermits er niet op elk sampletype evenveel getest wordt, wordt extra aandacht besteed aan de verschillen in test aantallen. Meer uitleg hierover volgt in de aparte sectie ‘Onderzoeksvraag’ en uiteraard in de ‘Materiaal en methode’. Ook wordt de interdisciplinariteit van het project nauwgezet in het oog gehouden, waardoor zowel informatici als biologen de kans hebben om een waardevolle bijdrage te kunnen leveren, alsook dat alle deelnemende studenten de opzet en uitvoering van het project kunnen begrijpen.

## **2 Onderzoeksvraag: Analyse van verboden middelen in verschillende sample collection methoden**

### **2.1 Oorspronkelijke Onderzoeksvraag**

Het oorspronkelijke onderzoek richtte zich op de detectie van illegale drugs in grote havens en de factoren die van invloed zijn op hun aanwezigheid. Het doel was om geografische verschillen in de aanwezigheid van illegale drugs te identificeren en te begrijpen waarom deze verschillen optreden.

## 2.2 Probleem met Beschikbare Data

Na het samenvoegen van annotaties en metadata, bleek er onvoldoende data beschikbaar te zijn over de havens. Dit resulteerde in een beperkt aantal overeenkomsten bij het uitvoeren van het Eclat algoritme, een methode om frequentiepatronen in datasets te identificeren. Hierdoor kon de oorspronkelijke onderzoeksvraag niet adequaat worden beantwoord.

## 2.3 Herziene Onderzoeksvraag

Als gevolg van de beperkte data over havens is de onderzoeksvraag aangepast naar: "In welke sample collection methoden worden de meeste verboden middelen aangetroffen?"

## 2.4 Nieuwe Focus

De herziene onderzoeksvraag richt zich op het analyseren van verschillende methoden voor het verzamelen van monsters en de abundantie van verboden middelen in die methoden. Daarnaast wordt gekeken naar het vermogen van deze methoden om verboden middelen te detecteren. De nieuwe focus omvat een breed scala aan sample collection methoden, zoals:

- Milieu-monsters
- Monsters van dieren
- Andere relevante bronnen

## 2.5 Doel van de Herziene Onderzoeksvraag

Het doel van de herziene onderzoeksvraag is om inzicht te krijgen in welke methoden het meest effectief zijn bij het identificeren van verboden middelen. Verder wordt onderzocht waarom er mogelijk variaties in detectiepercentages bestaan tussen de verschillende methoden. Dit kan leiden tot een beter begrip van de verspreiding en detectie van verboden middelen.

## 2.6 Belang van Flexibele Onderzoeksbenadering

Door de focus te verleggen van specifieke locaties naar diverse sample collection methoden, wordt een breder scala aan gegevensbronnen benut. Dit kan leiden tot een dieper inzicht in de verspreiding en detectie van verboden middelen. Het benadrukt ook het belang van een flexibele en aanpasbare onderzoeksbenadering, die reageert op de beschikbare gegevens en de behoefte om relevante vragen te beantwoorden binnen het kader van het onderzoek.

## 3 Materiaal en methode

### 3.1 Data verkrijgen

De spectrum annotaties zijn gedownload als tab-separated bestanden via de GNPS-website. Hiervan werden 8 verschillende bestanden gebruikt, met verschillende inputdata waarop de spectrumanalyse is uitgevoerd. Omwille van de grootte van deze bestanden werd een lazy loader gebruikt om deze te bekijken. De metadata van de verschillende experimenten waaruit de spectra afkomstig zijn, werden gedownload als tsv-bestand via de ReDU-website. Om meer informatie over de verschillende soorten medicijnen en drugs die geïdentificeerd worden via hun spectra in de experimenten te verzamelen, werd er een beroep gedaan op de DrugBank. Om de database hiervan te kunnen downloaden, werd er eerst een account aangemaakt dat goedgekeurd moest worden. Hierna kon het xml-bestand verkregen worden. Deze drie types van data werden als hyperlinks bezorgd door één van de begeleiders, Wout Bittremieux, via Slack.

### 3.2 Dataverwerking en bestanden joinen

Nu we de Annotaties, ReDU en drugsbank hebben moeten we deze aan elkaar kunnen linken. Dit deden we door joins uit te voeren tussen de verschillende databanken. Eerst hebben we de ReDU-database samengevoegd met elks van de 8 annotaties. Dit deden we door in beide te kijken naar hun afkomstige bestandsnaam. Zo hadden we 8 nieuwe bestanden die beide de kolumnen van de annotaties en van ReDU hadden. Na de join was het duidelijk dat niet van alles in de ReDU-databank er een annotatie was, maar ook dat er meerdere annotaties waren voor entries in de ReDU-databank. Dit zorgde ervoor dat de resulterende bestanden nog steeds groot waren. Daarom lieten we ze in aparte bestanden staan voor nu. We hebben nu een dataset met metadata, maar hebben de naam van de compounds nog niet. We kunnen deze verkrijgen door een join te doen met de drugsbank. In de annotaties staat de naam van de drug er niet bij, maar er staat een inchi-key bij. Deze zelfde inchi-key vinden we terug in de drugbank. Nu is de dataset wel klein genoeg om alles in één bestand te zetten. Iets wat opvalt is dat door deze joins heel veel data zijn verloren onderweg door de verschillen in de verschillende databanken. Stel dat deze in de toekomst nog zouden uitbreiden is het perfect mogelijk onze code opnieuw uit te voeren en meer resultaten te krijgen.

### 3.3 Berekeningen

Uit de ReDU-bestanden gefilterd op sampletype kunnen de totale aantallen samples gehaald worden. Het aantal verboden middelen die hierin worden teruggevonden kan bepaald worden via de gejoinde bestanden waarin alleen ‘illicit’ drugs worden opgenomen. Door het aantal verboden middelen te delen door de totale aantallen, worden de verhoudingen proportioneel weergegeven

zodat de categorieën onderling kunnen vergeleken worden. Dit wordt nog vermenigvuldigd met 100 om er een procentuele verhouding van te maken.

### 3.4 ECLAT

In ons onderzoek hebben we het ECLAT-algoritme (Equivalence Class Clustering and bottom-up Lattice Traversal) toegepast om correlaties te identificeren tussen verschillende categorieën, zoals de naam van de drugs en de gezondheidstoestand. De gekozen categorieën zijn:

- Name
- Country
- DOIDCommonName
- HealthStatus
- Lifestage
- NCBITaxonomy
- BiologicalSex
- HumanPopulationDensity
- Class

We hebben ECLAT uitgevoerd op 12 datasets. Deze datasets zijn verkregen door eerst het reduceren van de oorspronkelijke bestanden op basis van elk Sample-Type. Vervolgens hebben we elk gefilterd bestand samengevoegd met de 8 spectrumannotaties. De uiteindelijke dataset is verkregen door deze samengevoegde bestanden te combineren met de DrugBank-database, gefilterd op illegale drugs. Uit deze analyse hebben we interessante gegevens verkregen, die we presenteren in de resultaten.

## 4 Resultaten en discussie verschillende sample-types

### 4.1 Resultaten

Ten eerste is een opvallende vaststelling in dit onderzoek dat er geen matches zijn gevonden met 100% verboden middelen, het betreft in alle gevallen drugs die voor bepaalde toepassingen wel toegestaan zijn of zich ook nog in de experimentele fase bevinden. De grootste aantallen aan verboden middelen worden teruggevonden in stalen afgenomen van dieren (inclusief mens). Dit betreft 334 matches, waarvan er 213 gevonden zijn in vloeistoffen zoals bloed en 121 in weefsel. Het verboden middel dat het vaakst in dit type van stalen wordt teruggevonden is amfetamine, ook bekend als speed. Het op één na hoogste

aantal wordt teruggevonden in het sampletype van bacterieculturen. Dit gaat over 298 matches, allemaal met Methyleendioxymethamfetamine (MDMA) of ecstasy. In de andere categorieën van sampletypes gaat het over lagere aantallen. In drank, levensloze objecten en 'blank' extracties worden geen verboden middelen teruggevonden. Een overzicht van de absolute aantallen van verboden middelen in de spectradatabank wordt weergegeven in figuur 1.

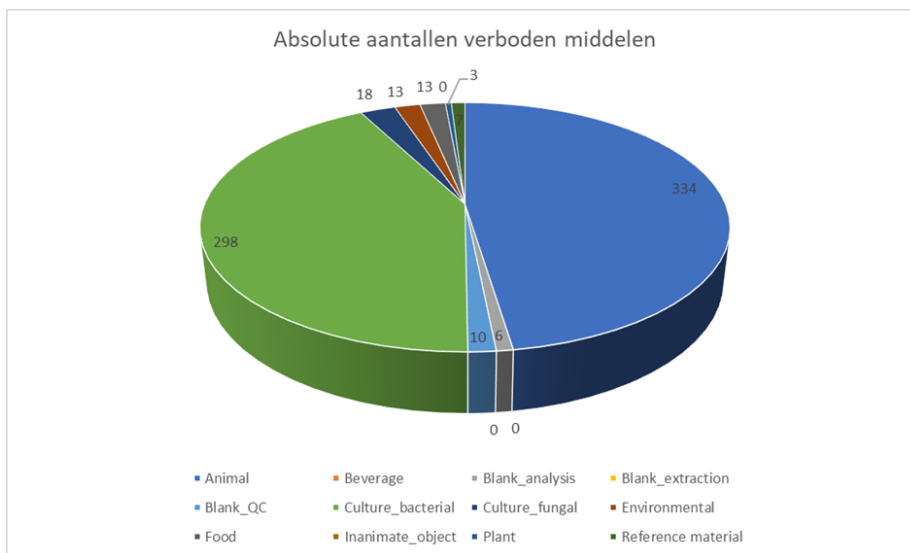


Figure 1: Taartdiagram van het aantal waarnemingen van verboden middelen in de verschillende sampletypes. De categorieën 'Animal' en 'Culture\_bacterial' zijn de koplopers met respectievelijk 334 en 298 waarnemingen. 'Culture\_fungal' staat in voor 18 matches, 'Environmental' en 'Food' elk voor 13, 'Blank\_QC' voor 10, 'Reference material' voor 7, 'Blank\_analysis' voor 6 en 'Plant' voor 3. De overige categorieën hebben geen waarnemingen

Er zijn natuurlijk niet van elk type evenveel stalen aanwezig in de database. Daarom is het om te onderzoeksvraag – in welk type staal zijn het meeste verboden middelen terug te vinden – te kunnen beantwoorden interessanter om naar de proportie van gevonden verboden middelen en het totaal aantal stalen van dat type te kijken. Deze zijn weergegeven in tabel 1, hierin worden ook samplesubtypes waar dit relevant is opgenomen. In figuur 2 worden deze aandelen van verboden middelen per sampletype visueel weergegeven. Hieruit komt de bacteriecultuur als duidelijke winnaar naar voren, met 9,55% van de stalen die een verboden middel bevatten. Daarna volgen omgevingsstalen uit de subcategorie stalen van opgelost organisch materiaal in het water genomen met 3,10% en stalen van referentiemateriaal uit de subcategorie chemische standaarden met 3,02%. Meer dan de helft van de (sub)categorieën bevat stalen waarin in minder dan 1% van de gevallen een verboden middel wordt gevonden.

<b>Samplotype</b>	<b>Proportie (%)</b>
Culture_bacterial	9,55
Environmental: dissolvedorgan- icmatter_water_saline	3,10
Reference material: refer- ence_material_chemicalstandard	3,02
Reference material	1,77
Animal: biofluid	1,66
Reference material: refer- ence_material_animalfeedersupplement	1,37
Environmental: culture_fungal	1,22
Blank_QC	1,15
Culture_fungal	1,14
Animal	1,09
Environmental	0,76
Animal: tissue	0,76
Food: food_source_plant	0,65
Food	44
Food: food_source_complex	0,24
Blank_analysis	23
Food: food_source_animal	0,12
Plant: plant_angiospermae	0,08
Plant	0,05
Plant: not specified	0,03
Beverage	0,00
Blank_extraction	0,00
Inanimate_object	0,00

Table 1: Overzicht van de procentuele proporties van verboden middelen aangetroffen in verschillende categorieën van stalen. Van sommige sampletypes zijn naast de hoofdcategorie ook subcategorieën opgenomen.

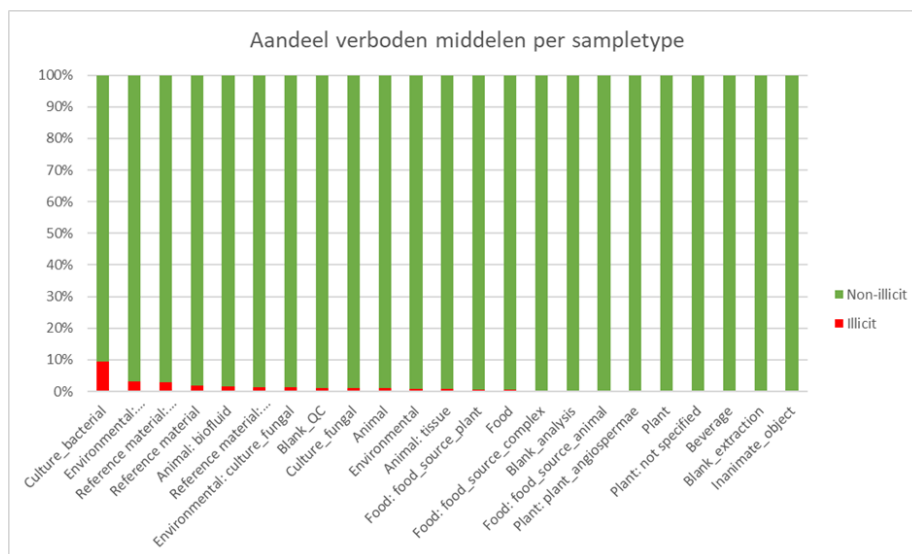


Figure 2: Stacked column met de proportie verboden middelen in vergelijking met het totale aantal samples. Voor de exacte waarden van de verschillende (sub)categorieën, zie tabel 1.

## 5 Resultatenanalyse

### 5.1 Hoogste Aandeel Verboden Middelen in Bacterieculturen

Uit de resultaten blijkt dat stalen afkomstig van bacterieculturen het hoogste aandeel aan verboden middelen bevatten, waarbij alle geconstateerde middelen hetzelfde type drug zijn, namelijk MDMA. Dit maakt het interessant om hier dieper op in te gaan.

### 5.2 MDMA in Bacterieculturen

MDMA is een synthetisch ontwikkelde drug met psychoactieve eigenschappen. Oorspronkelijk werd het door Merck pharmaceuticals ontwikkeld als eetlustremmer, maar het is nooit voor deze toepassing op de markt gebracht. Tegenwoordig wordt MDMA gezien als een potentieel hulpmiddel bij psychotherapie voor bijvoorbeeld de behandeling van PTSD, evenals als een partydrug (Connor, 2004). In de literatuur zijn geen bacteriën bekend die MDMA produceren. Ook is er geen studie die deze resultaten uit de databank ondersteunt, wat het moeilijk maakt om de hoge aanwezigheid van MDMA in de bacteriecultuurstalen te verklaren. Wel is bekend dat MDMA een onderdrukkend effect op het immuunsysteem heeft (Connor, 2004) en invloed kan hebben op leefgemeenschappen van bacteriën (Baslam et al., 2023). Hypothetisch gesproken zou het een nuttige stof kunnen zijn voor pathogene of concurrerende bacteriën om te produc-



eren. Door de schaarsheid aan informatie over de stalen en de context waarin ze zijn genomen (mogelijk is de MDMA door de onderzoekers zelf toegevoegd aan de bacterieculturen om het effect ervan te testen) is het moeilijk uitsluitsel te bieden over de reden dat de MDMA aanwezig is.

### 5.3 Verboden Middelen in Dierlijke Stalen

Een ander opvallend resultaat is dat dierlijke stalen in absolute aantallen de meeste verboden middelen bevatten (334 stalen versus 298 in bacterieculturen). In relatieve proporties is het echter de bacteriecultuur die het hoogste aandeel heeft (9,55% versus 1,09% in dierlijke stalen).

### 5.4 Vergelijking van Dierlijke Stalen en Bacterieculturen

Er bevinden zich meer stalen afkomstig van dieren in de databank (30,733) dan van bacterieculturen (3,120). Dit verschil in procentuele proporties kan veroorzaakt worden door verschillende factoren:

- De abundantie van verboden middelen in dierlijke stalen ligt mogelijk lager dan in bacterieculturen.
- Verboden middelen kunnen evenveel aanwezig zijn in beide categorieën, maar de detectie ervan is moeilijker in dieren dan in bacteriën.
- De relatief kleinere steekproefgrootte van bacterieculturen of het niet opnemen van de spectra van verboden middelen in de referentiedatabank bij stalen van dieren kan een vertekend beeld geven.

Zonder verdere informatie over de context waarin de stalen zijn afgenomen, is het moeilijk om uitspraken te doen over de verwachte hoeveelheid verboden middelen. Het is aannemelijk dat het moeilijker is om een drug op te sporen in een meercellig organisme met verschillende manieren om een staal af te nemen (urinstalen, uitwerpselen, bloed, huidstalen, enz.) dan in een eencellig organisme. Dit kan ook een verklaring zijn voor het lage aantal verboden middelen dat wordt teruggevonden in planten, waar natuurlijke drugs vandaan komen en er verschillende organen zijn om stalen van te nemen.

### 5.5 Verboden Middelen in Milieu- en Andere Stalen

De tweede categorie van sampletype in proporties is 'Environmental: dissolve-dorganicmatter\_water\_saline'. Dit was ook de categorie die voor de eerste onderzoeksvraag interessant was. In totaal zijn er maar 323 waarnemingen, waarvan slechts 10 illegaal zijn, wat weinig is om een analyse op te kunnen doen. Dit verklaart waarom het onderzoek naar de havens niet succesvol was.

## 5.6 Rioolwater en Drugs

De doorstroom van informatie uit onderzoeken met rioolwater en drugs naar de publieke database is duidelijk niet optimaal. Desalniettemin duidt de hoge percentuele proportie aan dat er potentie zit in deze onderzoeksvraag voor de toekomst, als het aantal totale stalen stijgt. Thai et al. (2014) merken op dat in rioolwater degradatie van sommige verboden middelen optreedt, wat een mogelijke belemmering voor deze onderzoeksmethode vormt.

## 5.7 Verboden Middelen in Voedsel en Referentiemateriaal

Ten slotte vallen de matches in voedsel en referentiemateriaal op, waar je vermoedelijk liever geen drugs in hebt. In plantaardig voedsel wordt een aantal keer codeïne gevonden, een opiaat dat uit de bolpapaver kan worden geëxtraheerd en als pijnstiller wordt gebruikt. De identiteit van de voedingsmiddelen kan niet uit de data gehaald worden, maar aangezien de bolpapaver giftig is, zal het vermoedelijk niet deze plant zijn en moet er eerder gedacht worden aan contaminatie.

## 5.8 Flunitrazepam in Chemische Standaarden

Flunitrazepam is een slaapmiddel dat meermaals voorkomt in chemische standaarden. Dit kan verklaard worden doordat flunitrazepam ook zelf als chemische standaard wordt gebruikt in toxicologisch onderzoek.

# 6 Individuele Eclat analyse

We gaan per sampletype een aantal interessante rules bespreken gekregen uit het eclat algoritme. Daarna gaan we vergelijken of deze rules blijven samenhangen doorheen de verschillende sampletypes. De splitsing vermindert de hoeveelheid data drastisch, en dus kan er bias zitten in de verkregen testresultaten.

## 6.1 Animal

Een van de grootste uitschieters is de relatie tussen Amphetamines en Urban omgeving. Ze hebben een support van 40% en een confidence van 72%. De lift is maar 1.12, maar toch duidt dit duidelijk op een overmatig gebruik van amfetamines in stedelijke gebieden. Dit kan wijzen op specifieke sociale of economische factoren die bijdragen aan het druggebruik in deze gebieden.

Chronic illness is iets wat over ter wereld voor lijkt te komen en niet specifiek is gebonden aan een regio of cultuur. In onze dataset ziet dit er heel anders uit. Wij vonden dat Chronic illness met een confidence van 80% gebonden is aan de VS. Dit geeft als resultaat een lift van 1.4. Aangezien de correlatie tussen de VS en chronic illness maar 20% is, betekent dit dat dit waarschijnlijk een onderrepresentatie is van chronic illness in andere gebieden. Als dat niet zo

is zegt dit duidelijk iets over de cultuur in de VS.

Er is een interessante correlatie tussen Phenethylamines en Male, met een support van maar liefst 34% en een confidence van 61%. De confidence van 61% lijkt op het eerste zicht niet zeer hoog, tot je ziet dat de confidence tussen Phenethylamines en Female maar 25% is. Er zijn dus veel gevallen waar het geslacht niet is bij verzameld. Met dit in ons achterhoofd is er dus een zeer grote correlatie tussen mannen en Phenethylamines, met een lift van 1.37 . Dit kan wijzen op een geslacht specifieke trend bij drugsgebruik en kan belangrijk zijn voor gerichte preventie en behandeling.

## 6.2 Culture Bacterial

Het is een dataset die volledig komt uit de Verenigde Staten. De hele dataset bestaat uit resultaten over Methylenedioxyethamphetamine. Omdat deze data gehaald werd uit bacterieculturen valt er weinig info uit de halen behalve dat op een of andere manier MDMA is gevonden. Dit kan door verschillende redenen zijn. Dit kan enerzijds komen door een Toxicologische studie over de reactie van bacterien op MDMA tot gevonden culturen in het milieu.

## 6.3 Culture Fungal

We zitten hier met hetzelfde probleem als bij Culture Bacterial. Een dataset uit de Verenigde Staten waarbij weinig info is waar ze vandaan komt. Aangezien we wel de info hebben dat de sample uit een rural gebied komt, brengt dit wel wat vragen op. MDMA komt niet voor bij fungi. De zowat enige verklaring kan zijn dat ze opgenomen zijn uit hun omgeving, bijvoorbeeld door het dumpen van drugsafval in dit gebied.

## 6.4 Food

Alle drugs gevonden in food behoren tot de familie van opioïde, waar in de dataset Codeïne het meeste voorkomt. Dit is zeer opmerkelijk aangezien er van nature in geen enkele voeding opioïden voorkomen.

## 6.5 Totaal

Er is een associatie tussen 'Urban' en 'Benzene and substituted derivatives' met een support van 0.264, een confidence van 0.762, en een lift van 1.995 . Dit suggereert een sterke relatie tussen stedelijke gebieden en de aanwezigheid van benzeen en afgeleide stoffen. Dit kan duiden op significante milieuvervuiling of industrieel gebruik van deze chemische stoffen in stedelijke omgevingen. Deze inzichten zijn van belang voor milieubeleidsvorming en volksgezondheidsinitiatieven gericht op het verminderen van blootstelling aan potentieel schadelijke

stoffen in dichtbevolkte gebieden.

De associatie tussen 'Urban' en 'Phenethylamines' met een support van 0.217, een confidence van 0.627, en een lift van 2.123 laat zien dat er een sterke relatie bestaat tussen stedelijke gebieden en het gebruik van fenethylaminen. Dit kan wijzen op een culturele of sociale voorkeur voor deze stoffen in dichtbevolkte gebieden. Het onderzoek naar dit fenomeen zou waardevolle inzichten kunnen bieden voor beleidsmakers en gezondheidsinstanties om gerichte interventies te ontwikkelen die zich richten op drugspreventie en -educatie in stedelijke gemeenschappen.

De hoge lift en confidence van de associatie tussen 'United Kingdom' en 'acute illness', met een lift van 8.864 en een confidence van 0.938, duiden op een significante correlatie tussen acute ziekten en de United Kingdom. Deze gegevens zijn bijzonder belangrijk voor gezondheidszorgplanning in deze regio, omdat ze helpen identificeren waar en hoe medische middelen mogelijk geconcentreerd moeten worden om effectief te reageren op hoge incidenties van acute aandoeningen.

## 7 Illicit Drugs op wereldkaart

Dit zijn de meest gebruikte drugs voor alle landen in de dataset, gebaseerd op de data in de dataset.

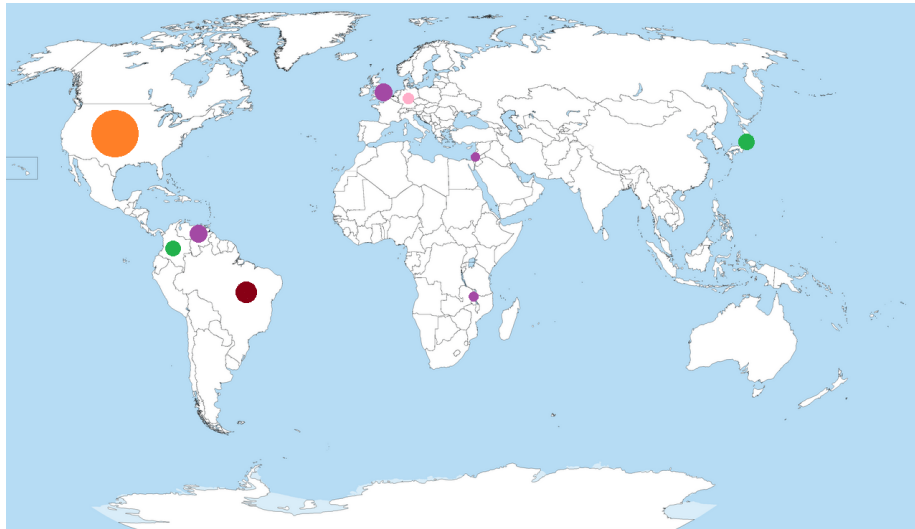


Figure 2: Wereldkaart met daarop de meestgebruikte drug volgens onze dataset

**MDMA:** Het is een recreatieve drug, vooral populair in de VS, zoals zichtbaar op de kaart, en wordt meestal gebruikt in een feestelijke setting. Door de analyse van elke sampletype individueel weten we dat we dit resultaat krijgen door een studie van bacteriële culturen en dus weten we de oorsprong hiervan niet echt.

**Amphetamine:** Dit is een stimulant voor bepaalde aandoeningen zoals ADHD en heeft naast dit ook recreatief gebruik.

**Androstenedione:** Een steroïde, die illicit is maar in landen met minder strenge regels rond steroïden populair kan zijn.

**Flunitrazepam:** Het is een drug die als bijnaam de "date rape drug" wordt genoemd. Gelukkig heeft het ook een medisch gebruik tegen slapeloosheid, waar we hopen dat de data vandaan komt.

**Etryptamine:** Het is een stimulerende drug die ontwikkeld werd als een antidepressivum, maar nooit op grote schaal werd geaccepteerd.

## 8 Conclusie

In bijna alle sampletypes komen verboden middelen voor. In absolute cijfers is dat het meest in de stalen afkomstig van dieren, in procentuele verhouding t.o.v. het totaal aantal afgenomen stalen is dit het geval voor de stalen van bacterieculturen. Er wordt gehypothetiseerd dat dit verschil te wijten zou kunnen zijn aan verschillen in het detectievermogen van verboden middelen in de verschillende sampletypes, maar ook alternatieve redenen worden niet uitgesloten. Uit de ECLAT-analyse komen enkele interessante links naar boven, zoals een overmatig gebruik van amfetamines in stedelijke context alsook een sterke relatie van benzeen(afgeleiden) hiermee. Tussen de UK en acute ziektebeelden wordt een link gemaakt door het algoritme, terwijl dit voor de VS net de chronische ziekten zijn. Deze resultaten kunnen nuttig zijn om mee te nemen in beleid rond volksgezondheid.

## 9 Individuele bijdragen

**Informatici: Elias, Philippe en Yassine**

**Bioloog: Amber**

**Elias** heeft zich beziggehouden met het voorbereidende werk en praktische zaken zoals een account aanmaken bij DrugBank om de benodigde data te verkrijgen, het opzetten van een GitHub voor dit project, etc. Hij heeft de parser voor de ReDU-database geschreven, alsook de gegeven parser voor DrugBank aangepast om specifiek verboden drugs eruit te filteren.

**Yassine** heeft het script geschreven om latitude en longitude coördinaten om te zetten in stadsnamen (nog voor de eerste onderzoeksvraag). Hij heeft de ReDU-data ook gefilterd op de verschillende sampletypes.

**Philippe** voerde de handelingen op zich uit: de verschillende bestanden joinen en het algoritme laten lopen (samen met **Yassine** om het wat meer vooruit te laten gaan). Als er hierbij problemen opdoken, gaf hij deze door aan de groep. Ook heeft hij nog verdere aanpassingen, die op dat moment nodig bleken, aan de code van Elias gemaakt.

**Amber** heeft de berekeningen uitgevoerd, plus de grafieken opgesteld met deze data. **Yassine, Elias en Amber** hebben tijdens deze fase van het proces gecoördineerd wat er moest gebeuren. Een aantal voorbeelden hiervan zijn het opstellen van een tijdsschema of het bepalen welke kolommen nuttige info bevatten en behouden moeten blijven en welke verwijderd mogen worden.

Er werden ook regelmatig overlegmomenten met de hele groep ingepland, waarbij iedereen actief deelnam om oplossingen voor opgedoken problemen aan te reiken. In de tussentijd werd ook overlegd in een Discord groep.

Voor het schrijven van de paper heeft **Amber** de meer theoretische uitleg over massaspectrometrie en metabolieten voor haar rekening genomen, terwijl de informatici de computationele methode schreven. **Elias** schreef de onderzoeksvraag uit en schreef de paper uit in latex, **Amber** het deel ‘Resultaten en discussie verschillende sampletypes’ en **Philippe** de analyse van de ECLAT-resultaten en de methode werd geschreven door **Philippe** en **Yassine**.

Alle **Code** en **Data** staat in onze **Github**: <https://github.com/elias-2060/Metabolomics>

## Literatuur

Baslam, A., Aitbaba, A., Hanchi, A.L., Tazart, Z., Aboufatima, R., Soraa, N., Ait-El-Mokhtar, M., Boussaa, S., Baslam, M. en Chait, A. (2023). Modulation of Gut Microbiome in Ecstasy/MDMA-Induced Behavioral and Biochemical Impairment in Rats and Potential of Post-Treatment with Anacyclus pyrethrum L. Aqueous Extract to Mitigate Adverse Effects. *International Journal of Molecular Sciences* 24, 9086.

Connor, T.J. (2004). Methylenedioxymethamphetamine (MDMA, ‘Ecstasy’): a stressor on the immune system. *Immunology* 111, 357-367.

Thai, P.K., Jiang, G., Gernjak, W., Yuan, Z., Lai, F.Y. en Mueller, J.F. (2014). Effects of sewer conditions on the degradation of selected illicit drugs residues in wastewater. *Water Research* 48, 538-547.